

Nucleotidsequenz der
konstanten Region der kappa-Ketten von
Maus-Immunglobulinen

Inaugural-Dissertation
zur Erlangung der Doktorwürde in der gesamten Medizin
an der Ludwig Maximilians-Universität München

Peter Neumaier

*Diss. München
1984*



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41548306_0087

Aus dem Institut für Physiologische Chemie, Physikalische
Biochemie und Zellbiologie der Universität München

Vorstände: Prof. Dr.rer.nat. Dr.med.h.c. Th. Bücher,
Prof. Dr.rer.nat. M. Klingenberg, Prof. Dr.med. F. Miller,
Prof. Dr.rer.nat. H.G. Zachau

Nucleotidsequenz der konstanten Region der kappa-Ketten von
Maus-Immunglobulinen

Inaugural - Dissertation
zur Erlangung der Doktorwürde in der gesamten Medizin
an der Ludwig Maximilians-Universität München

vorgelegt von
Peter Neumaier
aus
München
1981

Mit Genehmigung der Medizinischen Fakultät der
Universität München

Berichterstatter: Professor Dr. H.G. Zachau

Dekan: Professor Dr. W. Spann

Tag der mündlichen Prüfung: 19.1.1984

Die vorliegende Arbeit entstand am Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität München in der Zeit von April 1979 bis November 1980.

Herrn Prof. Dr. H.G. Zachau danke ich dafür, daß ich diese Arbeit unter seiner Anleitung anfertigen durfte. Mein besonderer Dank gilt Herrn Dr. W. Altenburger für seine Unterstützung und für zahlreiche Diskussionen während der Durchführung der Arbeit.

Meinen Eltern

Inhaltsverzeichnis

	Seite
1. Einleitung und Problemstellung	1
2. Ergebnisse	6
2.1. Ablauf der basenspezifischen DNA-Spaltungen	6
2.2. Interpretierbarkeit der Sequenzgele	11
2.3. Sequenzierungsstrategie.....	14
2.4. Primärstruktur des sequenzierten DNA-Abschnittes ..	15
3. Sequenzauswertung und Diskussion	19
3.1. Vergleich mit bisher veröffentlichten Sequenzen für die C _K -Region	19
3.2. Basenzusammensetzung	20
3.3. Dinucleotid-Häufigkeiten	22
3.4. Verwendung von Codons	25
3.5. Wiederholte, punktsymmetrische und spiegelsymmetri- sche Sequenzen	28
4. Materialien und Methoden	34
4.1. Materialien	34
4.1.1 Chemikalien und Reaktionsgefäße	34
4.1.2 Enzyme	34
4.1.3 Nucleinsäuren	34
4.1.4 Chromatographieplatten und Elektrophorese- apparaturen	35
4.1.5 Photographien und Autoradiographien	35
4.1.6 Computer	35
4.2. Methoden	35
4.2.1 Restriktionsspaltungen	35
4.2.2 Präparative Polyacrylamidgele und Elution von DNA .	36
4.2.3 Fällern, Umfällern und Waschen der DNA	36
4.2.4 Entfernung der 5'-ständigen Phosphatgruppen von DNA-Doppelsträngen	37

4.2. 5	[γ - ³² P]ATP-Synthese	37
4.2. 6	Markierung der 5'-Enden von DNA-Doppelsträngen mit ³² P	38
4.2. 7	DEAE-Cellulose-Säulen	39
4.2. 8	Basenmodifikationen	39
4.2. 9	Entfernung der modifizierten Basen und Spaltung der DNA-Stränge durch Piperidin	41
4.2.10	Denaturierende Polyacrylamidgele	41
4.2.11	Autoradiographie der Gele	42
4.2.12	Sequenzauswertung	42
5.	Zusammenfassung	48
6.	Literaturverzeichnis	50
7.	Abkürzungsverzeichnis	55
	Lebenslauf	

1. Einleitung und Problemstellung

Wirbeltiere können in den Organismus eingedrungene Krankheitserreger mit Hilfe von Antikörpern, die von differenzierten B-Lymphozyten (Plasmazellen) synthetisiert und sezerniert werden, binden und eliminieren. Die Fraktion der Immunglobuline im Blutplasma besteht aus einer äußerst heterogenen Population von Antikörpermolekülen, deren gemeinsame Grundstruktur von je zwei schweren und zwei leichten Polypeptidketten (H- und L-Ketten, 17) gebildet wird. Eine Plasmazelle produziert normalerweise nur einen Typ homogener Immunglobuline (Bibliographie in 67). Die Aufklärung der Mechanismen, die zur Antikörpervielfalt führen, wird durch multiple Myelome erleichtert. Diese Tumoren sind aus einer neoplastischen Transformation einer Plasmazelle entstandene Zellklone und produzieren in den meisten Fällen für biochemische Untersuchungen ausreichende Mengen homogener Proteine, die Antikörpern oder Teilen von Antikörpern entsprechen (18, 53). Als vorteilhaft erwies sich dabei die Verwendung von Mäusen aus Inzuchtstämmen (BALB/c, NZB), da bei ihnen multiple Myelome induziert (51) und durch Transplantation übertragen werden können (50).

Die Fähigkeit eines Antikörpers, eine bestimmte antigene Determinante spezifisch zu binden, wird von der Aminosäuresequenz seiner Polypeptidketten bestimmt (23, 83, 19). Sequenzanalysen zeigten, daß sich L-Ketten aus verschiedenen Antikörpern vor allem in ihren aminoterminalen Hälften (V-Regionen) unterscheiden (76, 26), wobei sich die Heterogenität auf drei hypervariable Regionen (84) konzentriert. Die

carboxylterminalen Hälften (C-Regionen) sind bezüglich ihrer Aminosäuresequenz konstant und können zur Definition von Typen und Subtypen (κ , λ_I und λ_{II} bei der Maus) von L-Ketten benutzt werden. Diese unterschiedliche Variabilität führte zur Annahme, daß verschiedene, voneinander getrennte Gene für V- und C-Regionen von L-Ketten codieren (16).

In den letzten Jahren entwickelte Techniken der in-vitro-Neukombination von Nucleinsäuren ermöglichten es, neben anderen eukaryontischen Genen auch solche, die für L-Ketten von BALB/c-Mäusen codieren, zu isolieren und durch Vermehrung in geeigneten Vektoren genaueren biochemischen Analysen zugänglich zu machen. Dabei stellte sich heraus, daß in Mauszellen, die keine Immunglobuline produzieren (z.B. Embryo, Leber, Spermien), die genetische Information für L-Ketten vom κ -Typ auf C-, J-, V- und L-Gensegmente verteilt ist. Etwa 2,4-4,0 kb in 5'-Richtung vom C_κ -Gensegment, das für die konstante Region von κ -Ketten codiert, liegen fünf voneinander durch jeweils etwa 300 Nucleotide getrennte J-Gensegmente. Jedes von ihnen kann für die Aminosäuren 96-108 der variablen Region von κ -Ketten codieren (38, 57). Die Anzahl der V_κ -Gensegmente, die für die Aminosäuren 1-95 der V_κ -Region codieren, wird auf etwa 200-1000 geschätzt (62). Sie liegen in einem unbekannten Abstand von den J-Gensegmenten und tragen auf ihrer 3'-Seite charakteristische Sequenzen, die komplementär zu Teilen der 5'-Seiten der J-Gensegmente sind (38, 57, 63). Jeweils etwa 100-200 Nucleotide vor den V_κ -Gensegmenten liegende Genfragmente codieren für hydrophobe Oligopeptide.

Diese an den Aminoterminus von κ -Ketten gebundenen Präsequenzen spielen wahrscheinlich bei der Sekretion von Antikörpern eine Rolle und werden dabei von dem sezernierten Polypeptid abgetrennt (42, 9). Die für die hydrophoben Präsequenzen codierenden Genfragmente können unterbrochen sein und direkt an V_{κ} -Gensegmente anschließen (2).

Ein aktives Gen für eine κ -Kette wird erst durch eine somatische Umordnung dieser Gensegmente geschaffen. Dabei wird ein Beitrag zur Antikörpervielfalt geleistet (79, 80), indem ein V_{κ} -Gensegment direkt an eines der fünf J-Gensegmente herangebracht wird. Bei dieser Rekombination fungieren möglicherweise die charakteristischen Sequenzen an den 3'-Seiten der V- bzw. den 5'-Seiten der J-Gensegmente als spezifische Erkennungssequenzen (38, 57, 63). Das aktive Gen ist wie viele andere eukaryontische und virale Gene (Bibliographie in 34) von intervenierenden Sequenzen oder Introns¹⁾ unterbrochen, also von DNA-Abschnitten, die zwar in hnRNA (heterogenous nuclear RNA) transkribiert, aber in einem "RNA-Spleißen" genannten Prozeß (Übersicht in 1) aus den mRNA-Vorläufern entfernt werden. Die DNA-Bereiche zwischen dem L- und dem V_{κ} -Gensegment und zwischen dem rekombinierten J- und dem C_{κ} -Gensegment bilden solche Introns. Die biologisch aktive mRNA für eine κ -Kette besteht aus einem nicht-translatierten Teil am 5'-Ende, einer für die hydrophobe Präsequenz, die V- und die C-Region codierenden Sequenz, einem nicht-translatierten Teil nach dem Terminatorcodon (3'UT) und einem am 3'-Ende angehefteten Poly-(A)-Segment.

1) "Intervenierende Sequenz" und "Intron" sind synonyme Begriffe. In der vorliegenden Arbeit werde ich "Intron" verwenden (siehe auch 20).

Mach et al. isolierten aus einem Myelom, das für MOPC 173 gehalten wurde, mRNA für κ -Ketten, retrotranskribierten sie in cDNA und klonierten das Retrotranscript in Plasmiden (Hybridplasmid K38, 35). Durch Restriktionsspaltungen des Plasmids K38 und elektrophoretische Auftrennung der dabei entstandenen Fragmente konnte unter anderem auch ein für den größten Teil der C-Region und 3'UT codierender DNA-Abschnitt isoliert werden; dieses Fragment wurde radioaktiv markiert und zur Identifizierung von C_{κ} -Gensegmenten gegen EcoRI-Spalstücke von DNA aus Leber- und aus Myelomgewebe hybridisiert (69, 71). Fragmente mit einem C_{κ} -Gensegment wurden in den Phagenvektor Charon 4A (6) eingebaut und in *Escherichia coli* amplifiziert (71, 70). Steinmetz et al. isolierten aus diesen Klonen durch Restriktionsspaltungen kleinere Fragmente mit V_{κ} -, J- und C_{κ} -Gensegmenten und subklonierten sie im Plasmid pBR322 (70 und M. Steinmetz, unveröffentlichte Daten). In dieser Versuchsreihe konnten bei dem verwendeten Myelom zwei umgelagerte V_{κ} -Gensegmente nachgewiesen werden (71, 70), von denen jedoch keines für die V-Region von κ -Ketten des MOPC 173 (61) oder eines anderen Mäusemyeloms mit bekannter Aminosäuresequenz codieren kann (2), weswegen der Tumor vorerst als "Myelom T" bezeichnet wurde (70).

In der vorliegenden Arbeit soll anhand von subklonierten Genomfragmenten aus Leber- und Tumorgewebe die Nucleotidsequenz des C_{κ} -Gensegments und des 3'-Teils des davor liegenden Introns bestimmt werden. Die Sequenz der vom C_{κ} -Gensegment transkribierten mRNA ist bereits veröffentlicht (24), auf

DNA-Ebene sind etwa 25% der im Rahmen der vorliegenden Dissertation zu bestimmenden Sequenz bekannt (38, 63). Der Vergleich von Sequenzen aus Leber und Myelom T ermöglicht eine Stellungnahme zu der Frage, ob somatische Mutationen bei der Umlagerung des C_{κ} -Gensegmentes stattgefunden haben. Ferner können im Rahmen der vorliegenden Arbeit mit Hilfe der Sequenzauswertung Beiträge zur Diskussion über die Evolution des C_{κ} -Gensegments und über möglicherweise funktionell wichtige Bereiche in der DNA-Primärstruktur geleistet werden.

2. Ergebnisse

2.1. Ablauf der basenspezifischen DNA-Spaltungen

Bei Anwendung der Methode von Maxam und Gilbert (39, 40) zur Aufklärung der Primärstruktur (Nucleotidsequenz) von DNA werden relativ große Mengen (etwa 10µg) homogener DNA-Fragmente, die an einem Ende radioaktiv markiert sind, in drei Teilreaktionen durch basenspezifische Spaltungen chemisch abgebaut. Die am Aufbau der DNA beteiligten Nucleotide werden auf chemischen Wege basenspezifisch modifiziert. Piperidin entfernt die modifizierten Basen aus der DNA und katalysiert die Eliminierung der Phosphatgruppen von den Desoxyribosen, an die die modifizierten Basen gebunden waren (39, 40). Diese Teilreaktionen sind mit Ausnahme der Purin-spezifischen Reaktion in Abb.1a-d schematisch dargestellt. Beim Purin-spezifischen Abbau werden die Glycosidbindungen von Purinbasen, die schon unter normalen Bedingungen labiler als die von Pyrimidinbasen sind (10), durch Ameisensäure gespalten (11, 31).

Die Zeit- und Temperaturbedingungen der ersten Teilreaktion sind dabei so zu wählen, daß die Basenmodifikationen nur bei wenigen Prozent der in Frage kommenden Nucleotide ablaufen. Dadurch entstehen bei den nachfolgenden Spaltungen DNA-Fragmente unterschiedlicher Länge, die auf denaturierenden Polyacrylamid(PAA)-Gelen elektrophoretisch aufgetrennt werden. Die Autoradiographie dieser Gele erfaßt nur radioaktiv markierte Fragmente und ermöglicht die Bestimmung eines Teils ihrer Nucleotidsequenz durch Zuordnung der Banden auf dem Röntgenfilm zu bestimmten Basen (39, 40).

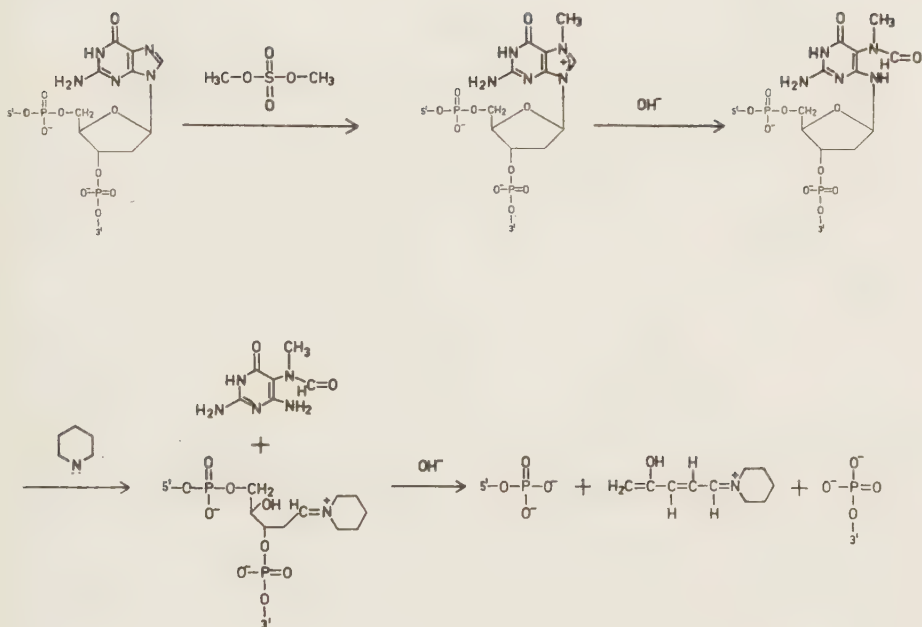


Abb. 1a: Mechanismus der basenspezifischen Modifizierung und DNA-Spaltung an Guanosin

Dimethylsulfat methyliert Desoxyguanosin an N7. Durch die zusätzliche Bindung etabliert sich zwischen N7 und N9 eine positive Ladung, wodurch die elektrophilen Eigenschaften des C8-Atoms (28) verstärkt werden. Unter basischen Bedingungen bricht der Imidazolring zwischen C8 und N9 auf (33, 40). Piperidin verdrängt das modifizierte Pyrimidin aus der DNA und katalysiert die Eliminierung beider Phosphate von der Desoxyribose. Nach (40).

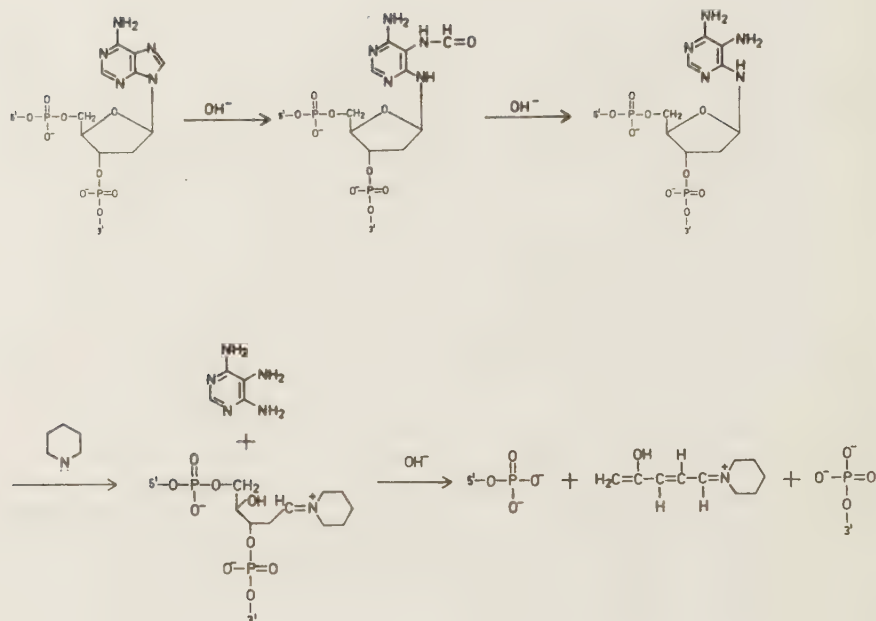


Abb. 1b: Wahrscheinlicher Mechanismus der basen-spezifischen Modifizierung und DNA-Spaltung an Adenosin

OH^- -Ionen greifen an dem in Purinen stark elektrophilen C8 (28) an, wodurch der Imidazolring zwischen C8 und N9 aufbricht. Unter stark basischen Bedingungen (1N NaOH, 90°C , siehe 4.2.8) reagiert das Produkt unter Abspaltung von Ameisensäure weiter zu 4,5-Diamino-6-amino(desoxyribosyl)pyrimidin (28). Piperidin verdrängt die modifizierte Base aus der DNA und katalysiert die Eliminierung beider Phosphate von der Desoxyribose (40). Über den Ablauf dieser Reaktion an Cytosin, das dadurch in geringerem Ausmaß modifiziert wird (40), sind keine Einzelheiten bekannt; es wird jedoch vermutet, daß sie mit dem Angriff von OH^- -Ionen am ebenfalls stark elektrophilen C4- (28, 29) oder C2-Atom (40) beginnt.

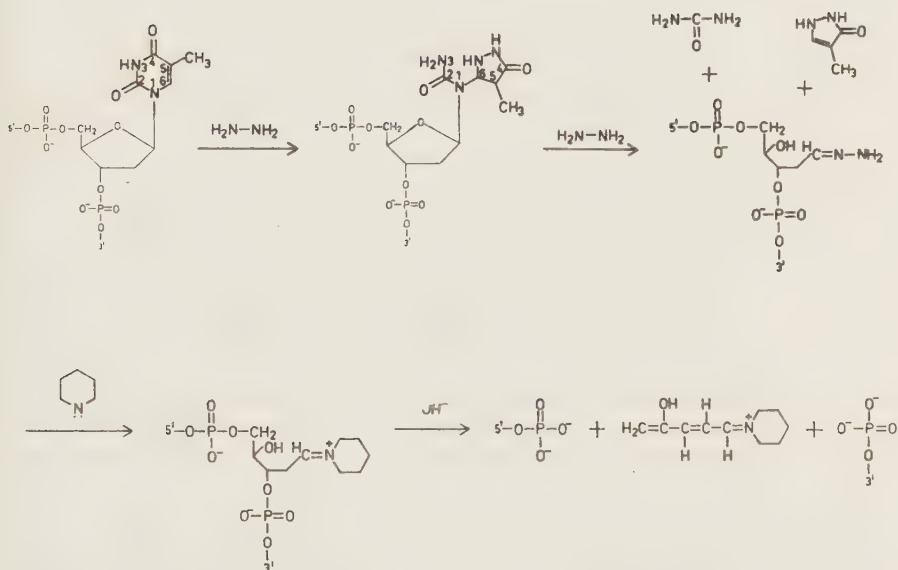


Abb. 1c: Mechanismus der basenspezifischen Modifizierung und DNA-Spaltung an Thymidin

Hydrazin öffnet den Pyrimidinring zwischen N3 und C4 und bildet mit C4-C5-C6 einen Pyrazolonring. Die Glycosidbindung wird hydrazinolytisch gelöst; als Produkte entstehen Harnstoff, 4-Methyl-pyrazolon-(5) und das Hydrazon des Desoxyribosylrestes (65, 40, Übersicht in 30). Nach der Piperidinspaltung des Hydrazons werden beide Phosphate vom Zuckerrest eliminiert. Nach (40).

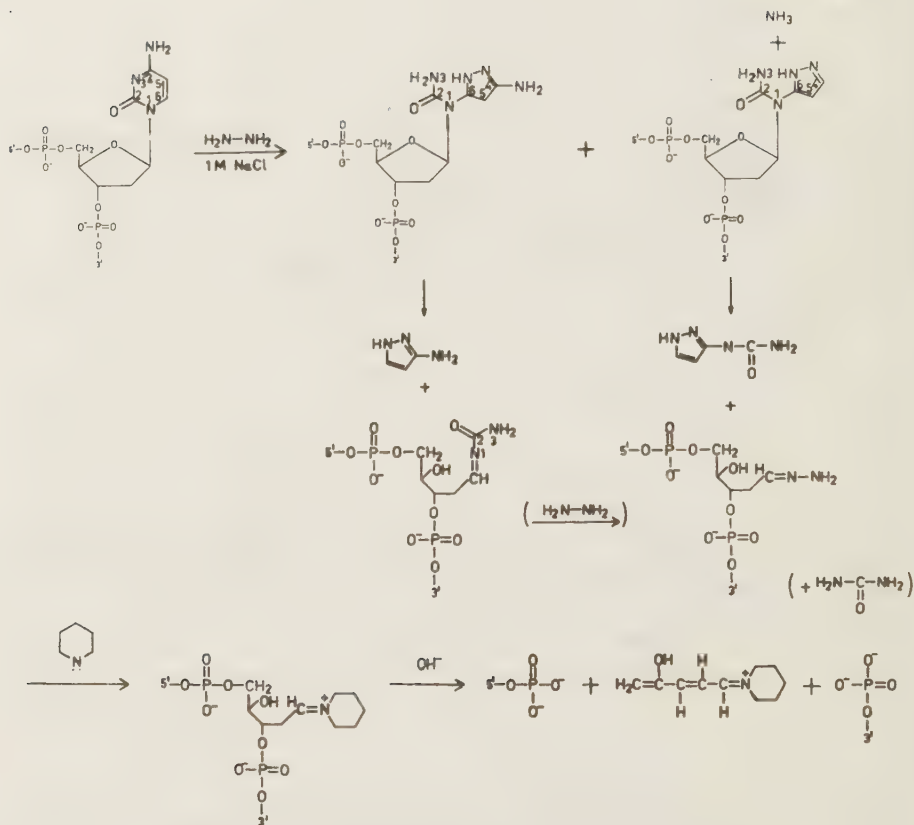


Abb. 1d: Mechanismus der basenspezifischen Modifizierung und DNA-Spaltung an Cytosin

Die Basenmodifizierung verläuft zunächst analog zu Abb. 1c (65, 40, Übersicht in 30), jedoch verlangsamt 1M NaCl die Modifizierung von Thymin erheblich (40); auch ohne Salz reagiert Thymin langsamer als Cytosin mit Hydrazin (30). Im Lauf der weiteren Reaktion mit Hydrazin entstehen neben Ammoniak, Harnstoff und dem meist als Hydrazone vorliegenden Desoxyribosylrest 3-Amino-pyrazol und 3-Ureido-pyrazol in gleichen Mengen (12). Piperidin reagiert unter Abspaltung der Hydrazone mit den Zuckerresten und katalysiert die Eliminierung beider Phosphate. Nach (12) und (40).

2.2. Interpretierbarkeit der Sequenzgele

Abhängig von der Dauer der Elektrophorese konnten Autoradiographien von 20%-PAA-Gelen maximal etwa bis zum 100., von 8%-PAA-Gelen etwa bis zum 260. Nucleotid, jeweils vom ^{32}P -markierten 5'-Ende an gerechnet, sicher gelesen werden; Tabelle 1 zeigt die Interpretierbarkeit von Sequenzgelen in Abhängigkeit von ihrer PAA-Konzentration und Laufzeit. Diese Angaben beziehen sich auf 0,25 mm dicke Gele und auf DNA-Fragmente, die vor den Basenmodifikationen über eine DEAE-Cellulose-Säule (siehe Teil Materialien und Methoden) gereinigt wurden. Die Auflösung ist bei Verwendung dickerer Gele und/oder ungereinigter DNA schlechter (zur Verwendung dünner Gele siehe auch 59). Etwa 2% der Banden in den gelesenen Sequenzen konnten keiner Base eindeutig zugeordnet werden; die betreffenden Nucleotide wurden durch wiederholte Sequenzierung und/oder Sequenzierung des Gegenstranges identifiziert. Die ersten drei Nucleotide am radioaktiv markierten Ende der DNA-Fragmente konnten in keinem Fall eindeutig bestimmt werden, die Banden waren unscharf, schwach, abgerundet und/oder keiner Base zuzuordnen (siehe auch Legende zu Fig. 1 in 44). Abb. 2 zeigt je ein Beispiel einer Autoradiographie eines 20%- und eines 8%-Sequenzgels.

a)

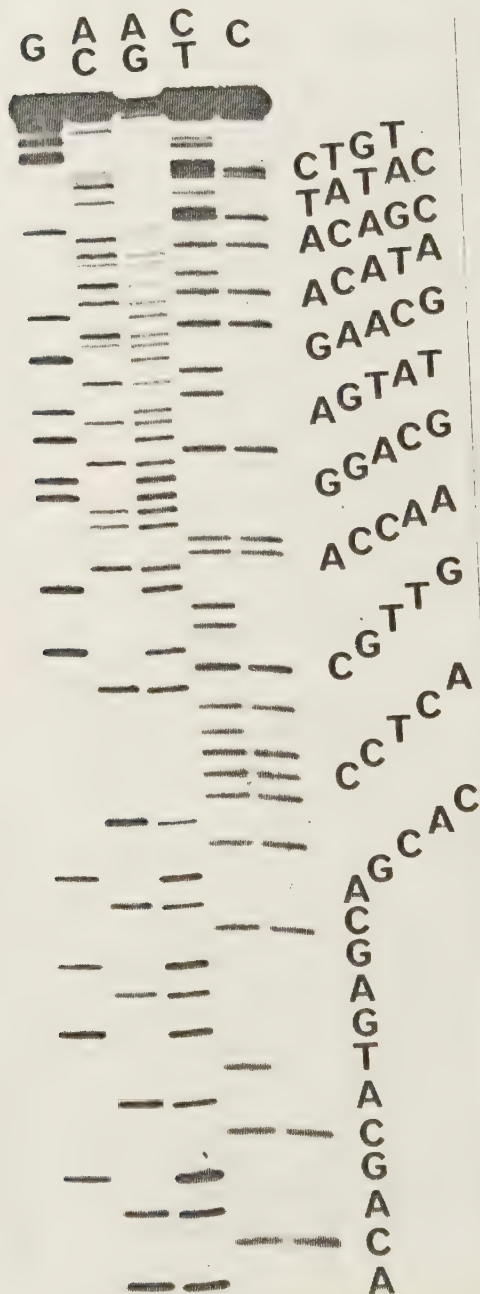


Abb. 2: Autoradiographien von Sequenzgelen

- a) Das an den 5'-Enden ^{32}P -markierte Sau3A/Sau96I Fragment (Position 657-817 in Abb. 4) wurde mit BspRI gespalten (Spaltung nach Position 750 in Abb. 4); die Spaltprodukte wurden über ein 5%-PAA-Gel isoliert. Das größere Fragment wurde den Abbaureaktionen nach Maxam und Gilbert (40) unterworfen. Die Autoradiographie eines denaturierenden 20%-PAA-Gels zeigt die Sequenz von Position 682 bis 746 in Abb. 4.
- b) Das an den 5'-Enden ^{32}P -markierte Sau96I/Sau3A Fragment (Position 9-656 in Abb. 4) wurde mit HinfI (Spaltstellen nach den Positionen 142, 179 und 270 in Abb. 4) gespalten. Nach Isolierung über ein 4%-PAA-Gel wurde das größte Fragment aus der HinfI-Spaltung nach Maxam und Gilbert (40) partiell abgebaut. Die Autoradiographie eines denaturierenden 8%-PAA-Gels zeigt die Nucleotidfolge des Gegenstranges der in Abb. 4 dargestellten Sequenz von Position 599 bis 510.

Gel %	Laufzeit	gelesene Nucleotide
20	1,5 h	4 - 40
20	6 h	25 - 80
8	3 h	65 - 150
8	4,5 h	100 - 220
8	6 h	120 - 250

Tabelle 1: Durchschnittliche Interpretierbarkeit von Sequenzgelen

Die Positionen der jeweils im Durchschnitt ersten und letzten sicher lesbaren Nucleotide ("gelesene Nucleotide"), vom radioaktiv markierten Ende der Fragmente an gerechnet, sind von der PAA-Konzentration ("Gel %") und von der Elektrophoresedauer ("Laufzeit") der Gele abhängig.

2.3. Sequenzierungsstrategie

Ausgangsmaterial der im Rahmen der vorliegenden Arbeit durchgeführten Sequenzanalysen waren subklonierte Genfragmente (70), die aus den Plasmiden isoliert und mit verschiedenen Restriktionsnucleasen gespalten worden waren (2). Die zwischen 100 und 800 bp langen Fragmente wurden an ihren 5'-Enden mit ^{32}P markiert und wieder mit Restriktionsnucleasen gespalten um jeweils zwei an nur einem Ende markierte Fragmente zu erhalten, die nach der Methode von Maxam und Gilbert (39, 40) sequenziert werden konnten. Ziel der in Abb. 3 dargestellten Sequenzierungsstrategie war es, durch mehrfache Sequenzierung und/oder Sequenzierung des Gegenstranges die ermittelte Nucleotidsequenz abzusichern. Die Verwendung von DNA aus Leber (Abb. 3b) und aus Myelomgewebe (Abb. 3c) sollte einen direkten Vergleich homologer Genfragmente aus immunologisch undifferenzierten und aus κ -Ketten produzierenden Zellen ermöglichen.

Da keine der Spaltstellen verschiedener Restriktionsnucleasen weiter als 180 bp von einer anderen entfernt war, ergaben sich bei der Sequenzierung dieses Bereiches keine Probleme. Eine Ausnahme bildet der Abschnitt zwischen den BspRI-Spaltstellen nach den Positionen 9 und 351 in Abb. 4, da es keine Restriktionsnuclease gibt, die darin nur einmal spaltet. Zur Bestimmung der Primärstruktur dieses DNA-Fragmentes wurde es nach der radioaktiven Markierung mit AluI partiell gespalten, d.h. durch eine Verkürzung der Inkubationszeit und eine Verringerung der eingesetzten Enzymmenge entstanden u.a. radioaktiv markierte Fragmente von Position 10 bis 296 und von Position 351 bis 33 in Abb. 4, die nach der o.g. Methode sequenziert werden konnten.

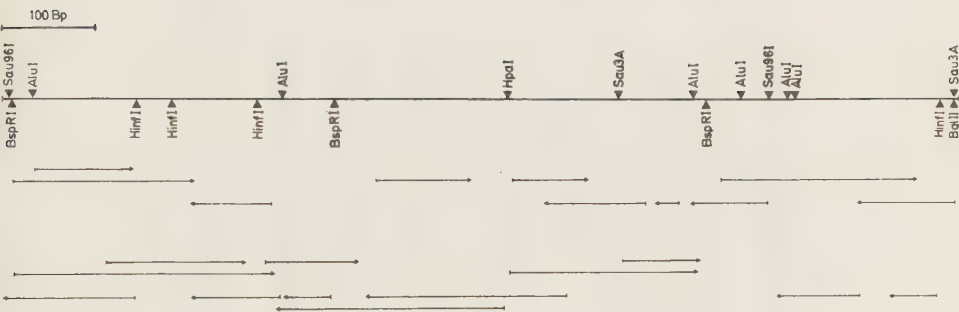


Abb. 3: Restriktionskarte und Sequenzierungsstrategie des sequenzierten DNA-Bereichs

- a) Restriktionskarte; nur die Spaltstellen der im Rahmen der vorliegenden Arbeit verwendeten Enzyme sind eingetragen.
- b) Sequenzierungsstrategie in DNA aus Lebergewebe. Die Pfeile geben jeweils Länge und Richtung der Teilsequenzen an.
- c) Sequenzierungsstrategie in DNA aus Myelom T.

2.4. Primärstruktur des sequenzierten DNA-Abschnittes

Abb. 4 zeigt die Nucleotidfolge des sequenzierten DNA-Bereichs von etwa 1 kb. Etwa 0,5 kb davon werden vom 3'-Teil des Introns zwischen den J-Gen-segmenten und dem C_K-Gensegment eingenommen, der Rest umfaßt das gesamte C_K-Gensegment mit dem nicht-translatierten Teil nach dem Terminatorcodon über die Poly-(A)-Anheftungsstelle hinaus.

Abschnitte, die sowohl in DNA aus Leber- als auch in DNA aus Myelomgewebe sequenziert wurden (siehe Abb. 3b und 3c), wiesen keine Unter-

Abb. 4: Legende siehe nächste Seite.

100 CTTAAGGCCCTGTCCATACAGTAGGTTAGCTTGGCTACACCAAGGAAGCCATACAGAGGCTAATATCAGAGTATCTTGGAAAGACAGGAGAAAA
200 AAAGCAGTTTTCTGCTTTACCTTATGTGCTTACCTTATGTGCTTACAGACTCCCAACATCAGGAAGTGTGAGATAAACTGGTCTGAACTCTCTGTCTGAAGCATGGA
300 GAAAAGAAATGTAGTTTCAGGGAAGAAAGCAATAGAAGGAAGCCCTGAGAATATCTTTCAAAGGGTCAGACTCAATTTACTTTCTAAAGAAGTAGCTAG
400 CTAAGGAATAACTTAGAAACAAGAATTGTATATATGTGTCATCCTGGCCCCATTGTTCCCTTATCTGTAGGGATAAGCGTGCCTTTTGTGTGCTGTG
500 AATAACATAACTGTTTACACATAATACACTGAAATGGAGCCCTTCCCTTGTTACTTCATACCATCCTCTGTGCTTCCCTTCCTCAGGGCTGATGCTGCA
600 ACTGTATCCATCTTCCCACCATCCAGTAGCAGATTAACTCTGGAGGGTGCCTCAAGTCGTGTGCTTCTTGAACAACCTTCTACCCCAAGACATCAATG
700 ThrValSerIlePheProProSerSerGluGlnLeuThrSerGluGlyAlaSerValValCysPheLeuAenAenPheTyrProLysAspIleAenV
800 AGTGGAAAGATTGATGGCAGTGAACGACAAAAATGGCGTCTGAAACAGTTGGACTGATCAGGACACCAAGACAGACCTACAGCATGAGCAGCACCCCT
900 ysTrpLysIleAaspGlySerGluArgIleAenGlnAenGlyValLeuAenSerTrpThrAaspGlnAaspSerLysAaspSerThrTyrSerMetSerSerThrLe
1000 GTTGACCAAGGACGAGTATGACGACATTAACAGCATTAACAGCATTAACAGCAGTGTGAGGGCACTCAAGACATCAACTTTCACCCATTGTCAAGAGCTTCAACAGGAAT
1100 rLeuThrLysAaspGluTrpGluArgHisAenSerTyrThrCysGluAlaThrHisLysThrSerThrSerProIleValLysSerPheAenArgAen
1200 TGTGTAGAGACAAGGTCCTGAGAGCGCCACCACCAAGCTCCCCAGCTCCATCCTATCTTCCCTTCTAAGGCTTGGAGGCTTCCCCACAACGACCTAC
1300 CysTer
1400 TGTGGGTGCTCCAAAGCTCCTCCCAACCTCCTCTCCCTTCTGCTTGTGCTTTTATCATGCTAATATTTGCAGAAAAATTATTCATAAA
1500 AGTCTTTGCACTTGAGATCT

Abb. 4: Primärstruktur des sequenzierten DNA-Abschnittes

Von 5' nach 3' ist der nicht-codierende, d.h. der mRNA homologe Strang dargestellt. Der für die C_K-Region codierende Bereich (Position 487-806) ist durch die den Codons entsprechenden Aminosäuren kenntlich gemacht. Ein Pfeil weist auf die nach dem Dinucleotid AG anzunehmende Spleißstelle (8) des Introns hin. Die am 3'-Ende eukaryontischer Gene hochkonservierte Sequenz AATAAA (52, Zusammenstellung in 3) ab Position 995 ist unterstrichen. Anheftungsstellen für Poly-(A) und damit Endpunkte von 3'UT können bei einem der beiden durch Pfeile gekennzeichneten Nucleotide an Position 1017 bzw. 1018 liegen.

schiede in ihrer Primärstruktur auf. Dieser Befund spricht gegen eine wesentliche Rolle von somatischen Mutationen in diesem Bereich bei der Umlagerung des C_K-Gensegments.

Eine nach der Nucleotidfolge zu erwartende Spaltstelle für die Restriktionsnuclease Sau96I (72) wird von diesem Enzym in DNA-Fragmenten, die in pBR322 subkloniert wurden, nicht erkannt. Bei der Vermehrung der DNA-Fragmente in *Escherichia coli* wurde ein C, das gleichzeitig Bestandteil der Erkennungssequenz für Sau96I und einer direkt daran anschließenden Erkennungssequenz für EcoRII ist, von der EcoRII-Methylase (7) in den Wirtsbakterien methyliert. Der betreffende DNA-Abschnitt ist in Abb. 5 dargestellt. Sau96I konnte an dieser Stelle nicht spalten, weil das Vorliegen eines m⁵C in der

Erkennungssequenz zur vollständigen Hemmung dieses Enzyms an dieser Stelle ausreicht (47, 46). Da m^5C durch Hydrazin nur sehr langsam modifiziert wird, (44, 41, 40), war in Autoradiographien von Sequenzgelen dieses methylierte Nucleotid durch Lücken in allen basenspezifischen Bahnen erkennbar.

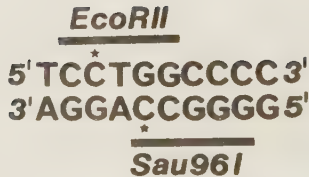


Abb. 5: Ausschnitt der DNA-Sequenz von Position 346 bis Position 355 in Abb. 4 mit einander überlappenden Erkennungssequenzen für Sau96I (72) und EcoRII (7); die EcoRII-Methylase methyliert die mit einem Stern gekennzeichneten Nucleotide (siehe Text).

3. Sequenzauswertung und Diskussion

3.1. Vergleich mit bisher veröffentlichten Sequenzen für die C -Region

In Arbeiten von Hamlyn et al. (24) und Max et al. (38) wurden Teile der Nucleotidfolge des in der vorliegenden Arbeit sequenzierten DNA-Abschnittes veröffentlicht. Hamlyn et al. sequenzierten die konstante und die 3'-nicht-translatierte Region der mRNA für κ -Ketten aus MOPC 21 (Position 487-1018 in Abb. 4), Max et al. DNA (Position 411-653 in Abb. 4). In den betreffenden Bereichen unterscheidet sich die Sequenz von Max et al. nicht, die von Hamlyn et al. nur an den 5'- und 3'-Enden von der im Rahmen dieser Dissertation ermittelten DNA-Sequenz. Diese Unterschiede sind in Abb. 6 dargestellt. Daraus wird auch

cDNA	3'	⁵³² GCCC	³¹⁷ ACTA....	¹⁰ G....	¹⁰ AAACGTGAAC.....	5'
mRNA	5'	¹ CGGC	²⁹⁶ CUGAU....	⁵²³ C....	⁵³³ UUUGCACUUGAAAAAA	3'
DNA		⁴⁸⁶ 5'AGGGG	⁷⁰¹ CTGAT....	¹⁰⁰⁰ C....	¹⁰¹⁸ TTTGCACTTGAGATCT	3'

Abb. 6: Vergleich der Sequenz von Hamlyn et al. ("cDNA" und "mRNA", 24) mit der in dieser Arbeit ermittelten Sequenz ("DNA"). Die Zahlen über den Nucleotiden beziehen sich auf die Position dieser Nucleotide in der jeweiligen Arbeit. An Position 317 der cDNA-Sequenz von Hamlyn et al. steht in (24) irrtümlich ein C. Durch Ausschneiden des Introns und durch Polyadenylierung des primären Transcripts im Verlauf des Reifungsprozesses von mRNA-Vorläufern zu biologisch aktiven mRNA-Molekülen entstehen die Unterschiede bei Position 486 und ab Position 1019 der DNA zu den entsprechenden Nucleotiden der mRNA. Nach der GT-AG-Regel (8) stellt das G an Position 486 das 3'-Ende des Introns dar.

ersichtlich, daß das erste A des Poly-(A)-Segmentes am 3'-Ende der mRNA von der DNA transkribiert werden kann und somit die Poly-(A)-Anheftungsstelle nicht genau lokalisierbar ist. Benoist et al. (3) halten dies für ein generelles Phänomen, weil man bei einem Vergleich der 3'-Enden anderer Gene ähnliche Verhältnisse findet.

3.2. Basenzusammensetzung

Eine Analyse der Basenzusammensetzung des nicht-codierenden Stranges, auf den sich alle folgenden Werte beziehen, ergab Unterschiede zwischen dem Intron, dem C_K-Gensegment und 3'UT (siehe Tabelle 2). Der sequenzierte Teil des Introns entspricht in seiner Basenzusammensetzung etwa der Gesamt-DNA aus Leberzellkernen von Mus musculus (66), die anderen Bereiche weichen z.T. stark davon ab; so sind z.B. die A+T-Anteile von C_K mit 50,75% und von 3'UT mit genau 50% niedriger als der A+T-Gehalt des sequenzierten Intronfragmentes (58,5%), Der A+T-Anteil des an spiegelsymmetrischen Sequenzen (s. 3.5) reichen, 100 Nucleotide umfassenden Abschnittes, der das Intron nach 3' begrenzt, liegt bei 62%; dieser Bereich enthält mit 64% auch einen hohen Pyrimidin-Anteil, was hauptsächlich auf einen hohen T-Gehalt (39%) zurückzuführen ist.

Die auffällige Häufigkeit von T im 3'-Teil des Introns paßt zu einem von Bina et al. (4) vorgeschlagenen Modell über das Herausschneiden von Introns aus mRNA-Vorläufern. Danach bilden das Poly-(A)-Segment, der 5'- und der 3'-Teil des Introns miteinander einen RNA-Tripelstrang, wobei das Poly-(A)-Segment und der 3'-Teil des

Introns antiparallel und durch Watson-Crick A-U-Paare miteinander verbunden sind und der 5'-Teil des Introns parallel zum Poly-(A)-Segment und durch andere als Watson-Crick-Bindungen daran gebunden ist. Auf diese Art können die Stellen, an denen das Intron ausgeschnitten und der so unterbrochene RNA-Strang wieder ligiert werden muß, in Nachbarschaft gebracht werden. Eine Voraussetzung für dieses Modell ist im Fall des in der vorliegenden Arbeit beschriebenen Intronfragments durch den hohen T-Gehalt des 3'-Teils erfüllt.

3'UT ist mit etwa 35% C- und 28% T-Gehalt (siehe Tabelle 2) reich an Pyrimidinen; hier fällt ein 39 Nucleotide langes Segment (Position 921-959 in Abb. 4) auf, das mit einer Ausnahme ausschließlich aus Pyrimidin-Nucleotiden besteht; darin kommt das Pentanucleotid CTCCT gehäuft vor.

Bereich	%A	%C	%G	%T	%A+T	%G+C	%A+G	%C+T	Länge
I	30,63	20,35	21,15	27,87	58,50	41,50	51,78	48,22	486
3'I	23,00	25,00	13,00	39,00	62,00	38,00	36,00	64,00	100
C _K +3'UT	27,07	29,70	19,55	23,68	50,57	49,25	46,62	53,38	532
C _K	30,63	26,25	22,50	20,63	51,26	48,75	53,13	46,88	320
3'UT	21,70	34,91	15,09	28,30	50,00	50,00	36,79	63,21	212
I+C _K +3'UT	28,81	25,14	20,33	25,72	54,53	45,47	49,14	50,86	1018
Gesamt	29,2	19,7	21,7	29,4	58,6	41,4	50,9	49,1	—

Tabelle 2: Basenzusammensetzung von verschiedenen Teilen der DNA-Sequenz in Abb. 4

I: sequenzierter Teil des Introns, Position 1-486;
 3'I: 3'-Teil des Introns, Position 387-486;
 C_K: für C_K-Region codierender Bereich, Position 487-806;
 3'UT: nichttranslatierter Bereich nach C_K, Position 807-1018; die Reihe "Gesamt" bezieht sich auf Leberzellkerne aus *Mus musculus* (66).

3.3. Dinucleotid-Häufigkeiten

Das Dinucleotid CG kommt bei der Maus wie auch bei vielen anderen Eukaryonten in der DNA aus Zellkernen seltener vor, als es aus der Basenzusammensetzung zu erwarten wäre (73, Datenzusammenstellung in 64). Da in eukaryontischer DNA fast ausschließlich die Sequenz CpG methyliert wird (zu $m^5\text{CpG}$, 15), schlug Salser (58) eine erhöhte Empfindlichkeit von $m^5\text{CpG}$ gegenüber Punktmutationen als Erklärung für das seltene Auftreten von CpG vor. Ein möglicher Mechanismus für eine solche Punktmutation besteht in der Desaminierung eines $m^5\text{C}$ zu T (60), wodurch das Dinucleotid CG durch TG (bzw. durch CA auf dem Gegenstrang nach der nächsten DNA-Replikation) ersetzt wird (5).

Aus den G- und C-Anteilen des sequenzierten DNA-Abschnitts ergibt sich für die Häufigkeit des Dinucleotids CG ein Erwartungswert (56) von $0,25 \cdot 0,20$, also von 5%. Tatsächlich tritt CG im sequenzierten DNA-Bereich 10 mal auf, das entspricht knapp 1%. Dabei ist die Verteilung zwischen dem Intron (einmal in etwa 0,5 kb; 0,2%) und dem C_{α} -Gensegment mit 3'UT (neunmal in 0,53 kb; 1,7%) ungleichmäßig. Um zu verlässlicheren Aussagen zu kommen, wurden für die in Tabelle 2 genannten Bereiche des sequenzierten DNA-Abschnitts die Häufigkeiten aller 16 Dinucleotide auf signifikante Abweichungen von den Erwartungswerten, die nach der Methode von Russell und Subak-Sharpe (56) berechnet wurden, untersucht, siehe Tabelle 3; die Werte für den 3'-Teil des Introns (3'I in Tabelle 2) blieben unberücksichtigt, weil dieser Bereich als zu kurz für eine solche Untersuchung angesehen wurde.

Als Signifikanzniveau wurde die doppelte Standardabweichung der Differenzen der Häufigkeit aller Dinucleotide (Daten aus 64) von ihren Erwartungswerten (berechnet nach Daten aus 66) in DNA aus Mäuseleberzellkernen gewählt; somit werden alle Differenzen, deren Absolutwert größer als 3,13 ist, als signifikant angesehen. Das Dinucleotid CG tritt in allen Bereichen signifikant selten auf, auch in C_K , weil hier wegen des vermehrten C+G-Gehaltes (siehe Tabelle 2) der Erwartungswert höher ist als im sequenzierten Teil des Introns. Die Dinucleotide TG und CA, deren vermehrtes Vorkommen nach Bird (5) den CG-Mangel ausgleicht, werden in allen Bereichen (von CA in 3'UT abgesehen) häufiger als erwartet gefunden, das Signifikanzniveau für diese beiden Dinucleotide wird jedoch nur in C_K überschritten. Diese Daten lassen unter Annahme der Hypothese von Bird (siehe oben) darauf schließen, daß im Laufe der Evolution des für die C_K -Region codierenden Genfragmentes CG-Dinucleotide methyliert und durch Punktmutationen zu TG bzw. CA aus diesem Segment eliminiert wurden. Für den sequenzierten Teil des Introns ist diese Elimination weder zu postulieren noch auszuschließen; zwar kommt CG hier signifikant selten vor, die Häufigkeiten für TG und CA überschreiten jedoch das Signifikanzniveau nicht und auch die Summe ihrer Abweichungen von den Erwartungswerten vermag den CG-Mangel nicht auszugleichen.

Jeweils knapp 2/3 der Dinucleotide TG (17 von 27) und CA (26 von 40) in C_K gehören zu zwei Codons, z.B. NNT-GNN. Wenn an einer solchen Position über den o.g. Mechanismus aus einem CG ein TG entsteht, hat dies keine Auswirkungen auf die Aminosäuresequenz der C-Region; findet diese Punktmutation aber auf dem Gegenstrang statt,

Tabelle 3: Legende siehe nächste Seite.

	G	I	G _K	3, UT	G _K +3, UT
AA	85 (+0,13)	48 (+0,62)	24 (-1,86)	13 (+1,45)	37 (-0,36)
AC	65 (-0,74)	24 (-1,00)	30 (+1,37)	11 (-2,36)	41 (-0,32)
AG	81 (+2,10)	43 (+2,35)	26 (+1,26)	12 (+2,41)	38 (+1,86)
AT	60 (-1,57)	33 (-1,91)	18 (-0,67)	9 (-1,88)	27 (-1,33)
CA	86 (+1,33)	31 (+0,44)	40 (+4,50)	15 (-0,46)	55 (+2,32)
CC	71 (+0,80)	22 (+0,72)	19 (-0,93)	30 (+2,03)	49 (+0,41)
CG	10 (-4,09)	1 (-3,98)	6 (-4,03)	3 (-3,85)	9 (-4,11)
CT	86 (+1,99)	41 (+2,86)	19 (+0,54)	26 (+2,44)	45 (+1,44)
GA	67 (+0,73)	32 (+0,08)	26 (+1,26)	9 (+0,99)	35 (+1,30)
GC	46 (-0,55)	19 (-0,27)	16 (-0,89)	11 (-0,06)	27 (-0,72)
GG	43 (+0,05)	24 (+0,37)	12 (-1,30)	6 (+0,57)	18 (-0,43)
GT	52 (-0,21)	28 (-0,37)	18 (+1,00)	6 (-1,43)	24 (-0,11)
TA	54 (-2,16)	37 (-1,08)	8 (-3,81)	9 (-1,88)	17 (-3,21)
TC	71 (+0,51)	30 (+0,59)	19 (+0,54)	22 (+0,55)	41 (+0,69)
TG	73 (+1,86)	35 (+1,10)	27 (+3,82)	11 (+0,94)	38 (+2,53)
TT	67 (-0,19)	37 (-0,55)	11 (-0,81)	18 (+0,52)	30 (+0,04)

Tabelle 3: Häufigkeit aller 16 möglichen Dinucleotide in verschiedenen Bereichen des sequenzierten DNA-Bereichs; G = Sequenz von Position 1-1018 in Abb. 4; I = sequenzierter Teil des Introns, Position 1-486 in Abb. 4; C_K = für die C_K-Region codierender Bereich, Position 487-806 in Abb. 4; 3'UT = nichttranslatierter Bereich nach C_K, Position 807-1018 in Abb. 4. Die Zahlen geben die absolute Häufigkeit der Dinucleotide im betreffenden Bereich an, die Werte in Klammern die Abweichungen der tatsächlichen Prozentanteile von den Erwartungswerten; hier sind alle Zahlen, deren Betrag größer als 3,13 ist, signifikant (siehe Text).

resultiert aus dem Ersatz eines mit G beginnenden Codons durch ein Triplet mit A an erster Position ein Aminosäureaustausch. Die relativ häufige Verwendung von Codons, die mit A beginnen (42 von 106, siehe Tabelle 4) und die Möglichkeit, daß 26 von ihnen auf diese Weise entstanden sein können, sprechen ebenfalls für eine Elimination von CG aus C_K durch Methylierung und Desaminierung. In den Bereichen C_K und C_K+3'UT ist ein signifikanter TA-Mangel festzustellen; eine plausible Erklärung dafür konnte nicht gefunden werden. Die Abweichungen aller anderen Dinucleotide von den Erwartungswerten sind nicht signifikant.

3.4. Verwendung von Codons

Miyata et al. führten für die Häufigkeit der Verwendung synonymen Triplets des degenerierten Codes einen Index f_a ein. f_a beschreibt das Verhältnis der Häufigkeit eines bestimmten Codons a (n_a) zur Häufigkeit aller Triplets ($\sum n_a$), die für die gleiche Aminosäure wie a codieren und mit a die ersten

beiden Nucleotide gemeinsam haben (43). Ihre Untersuchung aller bis dahin bekannten Sequenzdaten von Immunglobulin-mRNAs zeigte, daß für C-Regionen vornehmlich Codons codieren, deren drittes Nucleotid ein G oder C ist. Aus Tabelle 4 sind die absoluten und relativen Häufigkeiten der einzelnen Codons im C_{κ} -Gensegment und ihre f_a -Indices ersichtlich. Die Summe der f_a -Indices für Codons mit G oder C als letztem Nucleotid beträgt 12,366, für Codons mit A oder U als letztem Nucleotid 8,636. (Die nicht degenerierten Tripletts AUG für Met und UGG für Trp sind dabei nicht berücksichtigt.) Von 103 degenerierten Codonwörtern im C_{κ} -Gensegment enden 60, also 58,3% entweder mit G oder mit C, obwohl dieser DNA-Abschnitt mit 51,3% A+T-Gehalt (siehe Tabelle 2) nicht G+C-reich ist. Hasegawa et al. (25) finden bei der einzelsträngigen RNA des Phagen MS2 eine positive Korrelation zwischen der bevorzugten Verwendung von Codons, die mit G oder C enden, und RNA-Abschnitten, die an der Bildung von Sekundärstrukturen beteiligt sind, während in RNA-Abschnitten, die keine Sekundärstrukturen bilden, vornehmlich Codons mit A oder U an dritter Position gebraucht werden. Miyata et al. schließen aus diesem Befund und ihrer Untersuchung der Codon-Verwendung in Immunglobulinen auf eine ausgedehnte Sekundärstruktur in den für C-Regionen codierenden Teilen von Immunglobulin-mRNAs (43). Eine Suche nach Sequenzen mit punktsymmetrischen Elementen (s. 3.5) ergab bei dem im Rahmen dieser Dissertation sequenzierten C_{κ} -Gensegment der Maus keinen Anhaltspunkt für diese Annahme (siehe auch 24).

a)	b)	c)	d)	e)	a)	b)	c)	d)	e)
Arg	CGA	2	19	1	Gly	GGA	1	9	0,25
	CGC	0	0	0		GGC	2	19	0,5
	CGG	0	0	0		GGG	0	0	0
	CGU	0	0	0		GGU	1	9	0,25
Leu	AGA	0	0	0	Val	GUA	1	9	0,167
	AGG	1	9	1		GUC	4	38	0,667
	CUA	0	0	0		GUG	1	9	0,167
	CUC	1	9	0,5		GUU	0	0	0
Ser	CUG	1	9	0,5	Lys	AAA	2	19	0,286
	CUU	0	0	0		AAG	5	47	0,714
	UUA	1	9	0,333	Asn	AAC	5	47	0,625
	UUG	2	19	0,667		AAU	3	28	0,375
Thr	UCA	3	28	0,5	Gln	CAA	1	9	0,333
	UCC	2	19	0,333		CAG	2	19	0,667
	UCG	0	0	0	His	CAC	1	9	0,5
	UCU	1	9	0,167		CAU	1	9	0,5
Pro	AGC	7	66	0,7	Glu	GAA	2	19	0,333
	AGU	3	28	0,3		GAG	4	38	0,667
	ACA	2	19	0,182	Asp	GAC	4	38	0,571
	ACC	4	38	0,364		GAU	3	28	0,429
Ala	ACG	1	9	0,091	Tyr	UAC	2	19	0,5
	ACU	4	38	0,364		UAU	2	19	0,5
	CCA	3	28	0,6	Cys	UGC	1	9	0,333
	CCC	2	19	0,4		UGU	2	19	0,667
Ile	CCG	0	0	0	Phe	UUC	4	38	1
	CCU	0	0	0		UUU	0	0	0
	GCA	1	9	0,2	Met	AUA	0	0	0
	GCC	2	19	0,4		AUC	2	19	0,5
Trp	GGG	0	0	0		AUU	2	19	0,5
	GCU	2	19	0,4	Trp	AUG	1	9	-
						UGG	2	19	-

Tabelle 4:

Codon-Verwendung für die Aminosäuren 109-214 von κ -Ketten der Maus: a) Aminosäure; b) für a) codierende Basentriplets; c) absolute Häufigkeit von b); d) relative Häufigkeit von b) pro 1000 Aminosäuren; e) Häufigkeitsindex f_a für b) nach (43); $f_a = n_a / \sum n_a$ (siehe Text).

3.5. Wiederholte, punktsymmetrische und spiegelsymmetrische Sequenzen

Bei der Regulation der Genexpression und der Rekombination von Nucleinsäuren können Sequenzwiederholungen ("repeats", z.B. 5'...GATGAC...GATGAC...3'), Sequenzen mit punktsymmetrischen Elementen ("inverted repeats", z.B. 5'...GATGAC...GTCATC...3'), punktsymmetrische Sequenzen ("palindromes", z.B. 5'...GACGTC...3' und 3'...CTGCAG...5') und möglicherweise auch spiegelsymmetrische Sequenzen (z.B. 5'...GACCAG...3') eine Rolle spielen. Die von Thomas et al. und von Cavalier-Smith eingeführten und oben erwähnten englischen Begriffe (75, 82, 13) werden wegen ihres unterschiedlichen Gebrauchs in der Literatur hier nicht verwendet. Punktsymmetrische Elemente einer Sequenz können untereinander Wasserstoffbrücken ausbilden und so über die Formation eines doppelsträngigen Stammes mit einer daran hängenden einzelsträngigen Schleife die Sekundärstruktur von Nucleinsäuren beeinflussen (75, 82). Die Basenfolge einer spiegelsymmetrischen Sequenz lautet von 5' nach 3' gelesen ebenso wie von 3' nach 5'.

Der sequenzierte DNA-Bereich wurde mit Hilfe eines Computerprogrammes (siehe 4.2.12) nach solchen Mustern abgesucht; dabei wurden nur ununterbrochene Symmetrieelemente von sechs und mehr Nucleotiden berücksichtigt.

Im sequenzierten DNA-Abschnitt gibt es vier Typen von Sequenzen mit punktsymmetrischen Elementen (drei davon liegen im Intron; siehe Abb.7), die spiegelsymmetrische Sequenzen bzw. Teile davon enthalten und mehrfach vorkommen, also jeweils

Typ	Pos.	A	Pos.	B
I	45	<u>aAAGGAAGC</u>	358	tg <u>TTCTTa</u>
	238	<u>GAAGGAAGC</u>	443	c <u>CTTCCTT</u> g
			473	G <u>CTTCCTTC</u>
			952	ct <u>TTCTT</u> g

II	138	<u>TTcAGAcT</u>	172	<u>ctgGTCTGAA</u>
	266	<u>gTCAGAcT</u>	184	<u>tctGTCTGAA</u>

III	120	TtaccTTATGTG	406	aACATAActggtt
	332	<u>TGTATaTATGTG</u>	420	<u>CACATAATACAc</u>

IV	639	<u>GTCCTGA</u>	659	<u>CAGGACT</u>
	818	<u>GTCCTGA</u>		

Abb. 7: Vier Typen von spiegelsymmetrischen Sequenzen (A), die zu anderen Sequenzen (B) punktsymmetrisch angeordnet sind

"Pos." bezieht sich auf die Position des jeweils ersten angegebenen Nucleotids in Abb. 4. Die nicht in die jeweilige Punktsymmetrie passenden Basen sind durch Kleinschreibung gekennzeichnet; Spiegelsymmetrien sind unterstrichen.

mehr als eine Stamm-Schleifen-Struktur bilden können. Im Nonanucleotid GCTTCCTTC ab Position 473 in Abb. 4 (siehe auch Typ I in Abb. 7), das eine solche Struktur bilden kann, befindet sich ein Teil einer Sequenz, von der angenommen wird, daß sie beim RNA-Spleißen eine Rolle spielen könnte (34).

Die im sequenzierten DNA-Abschnitt liegenden spiegelsymmetrischen Sequenzen sind in Abb. 8 dargestellt. Sie kommen am 3'-Ende des Introns gehäuft vor; von 100 Nucleotiden (Position 387-486 in Abb. 8) liegen 60 in spiegelsymmetrischen Sequenzen, die sich paarweise überlappen. In 3'UT treten mehrere, relativ lange spiegelsymmetrische Sequenzen auf, von denen eine (ab Position 939 in Abb. 8) zusätzliche Symmetrie aufweist, da jede ihrer beiden Hälften für sich ebenfalls eine spiegelsymmetrische Sequenz darstellt. Es ist bis heute nicht geklärt, ob spiegelsymmetrische Sequenzen in Strukturgenen oder in Introns Funktionen ausüben und welcher Art solche Funktionen sein könnten.

Aus den Sequenzen mit punktsymmetrischen Elementen der Typen I, II und III in Abb. 7 läßt sich die in Abb. 9 gezeigte mögliche Sekundärstruktur des sequenzierten Teils des Introns konstruieren. Es ist nicht wahrscheinlich, daß eine solche Sekundärstruktur beim Ausschneiden des Introns aus dem mRNA-Vorläufer eine Rolle spielt, da große Teile von Introns ohne nachteilige Auswirkungen für das korrekte RNA-Spleißen deletiert werden können (74, 77). Gegen eine Beteiligung am RNA-Spleißen spricht auch die Mitwirkung der Sequenz 5'UCCUUC3' ab Position 476 in Abb. 9 am Aufbau der Sekundärstruktur, da u.a. diese Sequenz nach Meinung von Lerner et al. durch Ausbildung von Watson-Crick-Basenpaaren mit einem bestimmten Ribonucleoproteid im Zellkern zum korrekten Ausschneiden des Introns beitragen kann (34); in diesem Fall können keine Wasserstoffbrückenbindungen zu dem entsprechenden punktsymmetrischen Element gebildet werden. Eine mögliche Funktion dieser Sekundärstruktur, für deren Existenz es außer der DNA-Sequenz keine

Abb. 9: Hypothetische Sekundärstruktur eines Teils des Introns im mRNA-Vorläufer

Die Zahlen in den Schleifen geben jeweils deren Länge an, die Zahlen über den Nucleotiden beziehen sich auf die Position des jeweiligen Nucleotids in Abb. 4. Translatierte Codons sind durch die entsprechenden Aminosäuren und deren Position in κ -Ketten gekennzeichnet. Das Kreuz in Abb. 9a markiert die Stelle, an der die in Abb. 9b gezeigte Stamm-Schleifen-Struktur gebildet werden kann. Spiegelsymmetrische Sequenzen sind durch Striche gekennzeichnet.

Anhaltspunkte gibt, besteht in der Faltung eines ca. 350 Nucleotide langen Abschnitts des Introns zu einem relativ kompakten Gebilde. Interaktionen mit dem C_{κ} -Gensegment sind dabei nicht auszuschließen. Die Tatsache, daß alle am Aufbau dieser möglichen Sekundärstruktur beteiligten punktsymmetrischen Elemente spiegelsymmetrische Sequenzen bzw. Teile davon enthalten, konnte nicht befriedigend erklärt werden.

Anzahl und Lage von Sequenzen mit punktsymmetrischen Elementen im C_{κ} -Gensegment lassen nicht auf eine stabile Sekundärstruktur dieses Bereichs in der mRNA-schließen.

4. Materialien und Methoden

4.1. Materialien

4.1.1. Chemikalien und Reaktionsgefäße

Acrylamid, Harnstoff und Äthidiumbromid wurden von Serva, Heidelberg, bezogen, N,N'-Methylen-bisacrylamid von Biomol, Ilvesheim, Dithiotreitol von der Sigma Chemie GmbH, München. Hydrazin war von der Carl Roth KG, Karlsruhe, Spermidin-trihydrochlorid von der Fluka GmbH, Neu-Ulm, Xylen-cyanol von George T. Gutt Ltd., London. Boehringer, Mannheim, lieferte reduziertes Glutathion, NAD^+ , ATP und 3-Phosphoglycerat. (^{32}P)-Orthophosphat wurde von der New England Nuclear Corp., Dreieich, bezogen. DEAE-Cellulose (DE 52) war von Whatman Ltd., Springfield Mill, Kent, GB. Alle anderen, hier nicht aufgeführten Chemikalien stammen von Merck, Darmstadt.

1,5 ml- und 2 ml-Mikrogefäße wurden bei Sarstedt, Nümbrecht, 50 μl - und 100 μl -Glaskapillaren bei R.Brand, Wertheim, gekauft.

4.1.2. Enzyme

Glycerinaldehyd-3-Phosphat-Dehydrogenase und 3-Phosphoglycerat-Kinase zur Synthese von $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ waren von Boehringer, Mannheim, ebenso alkalische Phosphatase, T4-Polynucleotidkinase und die Restriktionsnucleasen HpaI und AluI. BspRI, HinfI, Sau3A und Sau96I stellte R.Streeck zur Verfügung.

4.1.3. Nucleinsäuren

Zu sequenzierende DNA wurde von M.Steinmetz kloniert (71, 70; siehe Einleitung). Lachs-DNA war von der Worthington Biochemical Corp., München, tRNA aus Hefe von Boehringer, Mannheim. Eichfragmente zur Größenbestimmung von DNA-Fragmenten unbekannter Länge stammten aus Laborbeständen.

4.1.4. Chromatographieplatten und Elektrophoreseapparaturen

Polyäthylenimin-Cellulose-Platten waren von Macherey-Nagel + Co., Düren, Gelkästen für präparative Polyacrylamid(PAA)-Plattengele (24 cm x 14 cm x 0,25 cm) und Elektrophoreseapparaturen für Plattengele zur Sequenzbestimmung (400 mm x 200 mm x 0,25 mm) wurden von unserer Werkstatt hergestellt.

4.1.5. Photographien und Autoradiographien

Präparative PAA-Gele wurden auf Gevatone Pan P230p (13 cm x 18 cm) von Agfa (entwickelt mit Agfa G5-P) photographiert, für Autoradiographien wurden Röntgenfilme von Kodak (X-Omat-R XR 5, 20 cm x 40 cm bzw. 18 cm x 24 cm, entwickelt mit Kodak D 19) verwendet.

4.1.6. Computer

Die Auswertung der ermittelten DNA-Sequenz erfolgte mit Hilfe eines 4052 Graphic Systems, erweitert durch einen 4642/1 Printer und einen 4662 Plotter der Firma Tektronix, Beaverton, Or., USA, und eines Teils der in 4.2.12 erläuterten Programme.

4.2. Methoden

4.2.1. Restriktionsspaltungen

Alle Restriktionsspaltungen wurden in 10 mM Tris-HCl (pH 7,5), 8 mM $MgCl_2$, 50 mM NaCl, 1 mM Dithiothreitol (70) bei 37°C durchgeführt. Durch Zugabe eines dem Volumen der Proben entsprechenden Volumens von 2,5 M Saccharose, 1 mM EDTA, 0,01% (wt/vol) Bromphenolblau wurden die Reaktionen gestoppt, die zum Auftragen auf Gele erforderliche Dichte erreicht und ein Vergleichsfarbstoff zur Festsetzung der Dauer der

Elektrophoresen zugemischt (27). Proben, die nicht sofort weiterverarbeitet wurden, wurden bei -40°C aufbewahrt.

4.2.2. Präparative Polyacrylamidgele und Elution von DNA

Lösungen für 4%- bzw. 5%- (wt/vol) PAA-Gele enthielten Acrylamid und N,N'-Methylen-bisacrylamid im Verhältnis 30:1, 89 mM Tris-Borat (pH 8,3), 2,5 mM EDTA (45) und 0,07% (wt/vol) Ammoniumperoxodisulfat. Zu den filtrierte und entgasten Lösungen wurden zum Start der Polymerisation 0,05% (vol/vol) TEMED gegeben, die Polymerisationsdauer betrug etwa 16 h. Nach einer Vorelektrophorese von etwa 2 h in 89 mM Tris-Borat (pH 8,3), 2,5 mM EDTA (45), wurden die DNA-Fragmente bei etwa 7 V/cm aufgetrennt. Nach der Elektrophorese wurden die DNA-Banden in den Gelen mit Äthidiumbromid (1 $\mu\text{g/ml}$) angefärbt (78), unter UV-Licht photographiert (48) und ausgeschnitten. Die Gelstücke wurden zu einer homogenen Masse zerrieben und die DNA mit 1 M NaCl daraus eluiert; die Eluate wurden zweimal durch Whatman 3MM-Papier filtriert (W.Fiers, persönliche Information).

4.2.3. Fällen, Umfällen und Waschen der DNA

Die Proben wurden mit dem dreifachen Volumen Äthanol vermischt, mindestens 15 min auf -70°C oder 16 h auf -20°C gekühlt und mindestens 10 min bei 12.000 g zentrifugiert. Die ausgefällte DNA wurde in 300 μl 0,3 M Natriumacetat, pH 5,8, gelöst, durch Zugabe von 1 ml Äthanol umgefällt und wie oben beschrieben gekühlt und abzentrifugiert (40). Daran schlossen sich zwei Waschschrirte mit je 1 ml 80% Äthanol an; die Proben wurden dabei nicht gekühlt und nur kurz (etwa 20 s) zentrifugiert. Nach etwa 30 min Lyophilysse wurden die Proben weiterverarbeitet.

4.2.4. Entfernung der 5'-ständigen Phosphatgruppen von DNA-Doppelsträngen

Zur enzymatischen Abspaltung der 5'-ständigen Phosphatgruppen (81) wurden zu der in 200 μ l 10 mM Tris-HCl, pH 8,0 gelösten DNA etwa 2 U alkalische Phosphatase aus Kälberdarm (14) gegeben und 30 min bei 60°C inkubiert. Durch zweifaches Ausschütteln mit je 200 μ l Phenol, gesättigt mit 50 mM Tris-HCl, pH 8,0, 1 mM EDTA, wurde das Enzym denaturiert und aus den Ansätzen extrahiert; darin verbliebenes Phenol wurde durch zweifache Ätherextraktion mit dem 2,5-fachen Probenvolumen Diäthyläther entfernt. Zu den Proben wurden 25 μ l 3 M Natriumacetat, pH 5,8, gegeben und die DNA wurde wie in 4.2.3 beschrieben gefällt und gewaschen (40).

4.2.5. [γ - 32 P]ATP-Synthese

Bei 4°C wurden 50 U Glycerinaldehyd-3-Phosphat-Dehydrogenase und 150 U 3-Phosphoglycerat-Kinase in Ammoniumsulfat-Suspension 5 min bei 12.000 g zentrifugiert und nach Abpipettieren des Überstandes zweimal mit je 100 μ l 3,2 M Ammoniumsulfat, 100 mM Tris-HCl, pH 8,0, 10 mM Mercaptoäthanol, 1 mM EDTA, 0,1 mM NAD^+ resuspendiert und 5 min bei 12.000 g zentrifugiert. Nach Abpipettieren des letzten Überstandes wurden die Enzyme in 37 μ l Wasser gelöst. Zu 20 μ l 500 mM Tris-HCl, pH 8,0, 50 mM MgCl_2 , 20 mM reduziertem Glutathion, 10 mM 3-Phosphoglycerat, 1 mM EDTA, 0,1 mM NAD^+ , 0,4 mM ATP wurden 25 mCi (entspricht etwa 2,5 nmol) (^{32}P)-Orthophosphat gegeben und mit Wasser auf 200 μ l aufgefüllt. Ein aliquoter Teil dieser Lösung wurde auf eine Polyäthylenimin-Cellulose-Platte aufgetragen. Durch Zugabe von 3 μ l der Enzymlösung wurde die Synthese von [γ - ^{32}P]ATP gestartet. Zur Beurteilung der Effektivität der Austauschreaktion wurde nach je 1, 3 und 5 min ein aliquoter Teil der Lösung

auf die Polyäthylenimin-Cellulose-Platte aufgetragen. Nach 6 min wurden die Enzyme durch Einfrieren bei -40°C denaturiert; danach wurde die Lösung mit $10\ \mu\text{l}$ $200\ \text{mM}$ EDTA, pH 7,4, $100\ \mu\text{l}$ Wasser und Äthanol auf $1\ \text{ml}$ aufgefüllt, auf 10 Portionen à $100\ \mu\text{l}$ verteilt und bei -40°C aufbewahrt. Nach der Chromatographie mit $0,75\ \text{M}$ Kaliumphosphat, pH 3,5, als Laufmittel wurde die Polyäthylenimin-Cellulose-Platte 10 s autoradiographiert und anhand des entwickelten Filmes die Effektivität der Austauschreaktion beurteilt. Das angegebene Verfahren entspricht der von Maxam und Gilbert (40) modifizierten Glynn-Chapell-Methode (21).

4.2.6. Markierung der 5'-Enden von DNA-Doppelsträngen mit ^{32}P

Lyophilisierte DNA, deren 5'-ständige Phosphatgruppen wie in 4.2.4 beschrieben entfernt worden waren, wurde in $10\ \mu\text{l}$ H_2O gelöst, nach Zugabe von $35\ \mu\text{l}$ $20\ \text{mM}$ Tris-HCl, pH 9,5, $1\ \text{mM}$ Spermidin, $0,1\ \text{mM}$ EDTA $4\ \text{min}$ bei 90°C denaturiert und sofort auf 0°C abgekühlt. Nach Zugabe von $5\ \mu\text{l}$ $500\ \text{mM}$ Tris-HCl, pH 9,5, $100\ \text{mM}$ MgCl_2 , $50\ \text{mM}$ Dithiothreitol, 50% (vol/vol) Glycerin wurde die Lösung in ein Mikrogefäß mit einer lyophilisierten Portion [γ - ^{32}P]ATP (siehe 4.2.5) umpipettiert und mit etwa $5\ \text{U}$ T4-Polynucleotidkinase (55) $30\ \text{min}$ bei 37°C inkubiert. Die Reaktion wurde durch Zugabe von $200\ \mu\text{l}$ $2,5\ \text{M}$ Ammoniumacetat, pH 5,8, abgebrochen und die DNA wie in 4.2.3 beschrieben gefällt, umgefällt und gewaschen (40). Anschließend wurde die DNA im Puffer für Restriktionsspaltungen (siehe 4.2.1) gelöst und $1\ \text{h}$ bei 68°C renaturiert. Nach Spaltung der DNA mit 2 - $4\ \text{U}$ einer Restriktionsnuclease (siehe 2.3) wurden die an nur noch einem 5'-Ende radioaktiv markierten Spaltprodukte wie in 4.2.2 beschrieben auf einem PAA-Gel aufgetrennt, daraus eluiert und wie in 4.2.3 beschrieben gefällt, umgefällt und gewaschen.

4.2.7. DEAE-Cellulose-Säulen

In mit Glaswolle verschlossene 1000 μ l Pipettenspitzen wurden 300 μ l DEAE-Cellulose-Suspension (68) in 1 M NaCl, 100 mM Tris-HCl, pH 7,8, 10 mM EDTA pipettiert und dreimal mit je 200 μ l Waschpuffer (0,1 M NaCl, 10 mM Tris-HCl, pH 7,8, 1 mM EDTA) gespült. Radioaktiv markierte DNA-Fragmente aus 4.2.6 wurden in 200 μ l Waschpuffer gelöst und auf die DEAE-Cellulose-Säulen aufgetragen. Die Säulen wurden mit 2 ml Waschpuffer, aufgeteilt in 10 Portionen à 200 μ l gewaschen. Nach dem letzten Waschschrift wurde die DNA mit 350 μ l 1 M NaCl, 10 mM Tris-HCl, pH 7,8, 1 mM EDTA eluiert und nach Zugabe von 5 μ l tRNA (5 mg/ml) und 5 μ l auf eine Länge von etwa 0,5 kb geschallter, phenolisierter und gegen Wasser dialysierter Lachs-DNA (1 mg/ml) (40) wie in 4.2.3 beschrieben gefällt, umgefällt und gewaschen.

4.2.8. Basenmodifikationen

Lyophilisierte DNA aus 4.2.7 wurde in 30 μ l Wasser gelöst und auf 5 Mikrogefäße à 6 μ l aufgeteilt.

Zur Abspaltung der Purinbasen wurde die DNA wieder lyophilisiert und in 15 μ l 66% (vol/vol) Ameisensäure, 2% (wt/vol) Diphenylamin, 1 mM EDTA gelöst (22, 11). Die Lösung wurde je nach Länge der zu sequenzierenden DNA-Fragmente für 2,5 min - 4 min bei Raumtemperatur geschüttelt. Die Zugabe von 45 μ l Wasser von 0°C brach die Reaktion ab. Diphenylamin und Ameisensäure wurden durch dreifache Ätherextraktion mit jeweils 500 μ l Diäthyläther aus den Proben entfernt; die DNA wurde lyophilisiert und wie in 4.2.9 beschrieben weiterverarbeitet.

Zur Methylierung von G wurden 200 μ l 50 mM Natriumcacodylat, pH 8,0, 10 mM $MgCl_2$, 1 mM EDTA und 1 μ l Dimethylsulfat zur DNA pipettiert; die Reaktion wurde nach 4 min - 15 min bei Raumtemperatur durch Zugabe von 50 μ l 1,5 M Natriumacetat, pH 7,0, 1 M Mercaptoäthanol, 100 μ g/ml tRNA, 0°C gestoppt (40). Die DNA wurde nach 4.2.3 gefällt, umgefällt und gewaschen.

Zur Modifikation von C und C + T wurden 15 μ l 5 M NaCl bzw. 15 μ l Wasser zur DNA gegeben und jeweils 30 μ l Hydrazin zugemischt. Die bei Raumtemperatur ablaufenden Reaktionen wurden nach 5 min - 15 min durch 200 μ l 0,3 M Natriumacetat, pH 5,8, 0,1 mM EDTA, 25 μ g/ml tRNA, 0°C abgebrochen (40). Die DNA wurde wie in 4.2.3 beschrieben gefällt, umgefällt und gewaschen.

Die Modifikation von A + C erfolgte durch 20 μ l 1,5 N NaOH, 1 mM EDTA in siliconisierten und über einer Flamme verschlossenen 100 μ l-Glas-kapillaren bei 90°C (39). Nach Ablauf der 10 min - 20 min betragenden Reaktionszeit wurden die Lösungen in den Kapillaren bei -70°C schockgefroren und nach dem Auftauen in Mikrogefäße mit 40 μ l 1 N Essigsäure umpipettiert. Die Kapillaren wurden mit je 100 μ l Wasser und 0,3 M Natriumacetat, pH 5,8, nachgespült. Die DNA wurde nach Zugabe von 1 μ l tRNA (5 mg/ml) wie in 4.2.3 beschrieben gefällt und gewaschen.

4.2.9. Entfernung der modifizierten Basen und Spaltung der DNA-Stränge durch Piperidin

Zu lyophilisierter und in 20 μ l Wasser gelöster DNA aus 4.2.8 wurden 5 μ l 2,5 M Piperidin gegeben; die Reaktionen liefen 30 min bei 90°C in siliconisierten und über einer Flamme verschlossenen 100 μ l-Glaskapillaren ab (39, 40). Zum Abbruch der Reaktion wurde die Lösung in den Kapillaren bei -70°C schockgefroren. Die DNA wurde in Mikrogefäße umpipettiert, lyophilisiert und zweimal in je 50 μ l Wasser gelöst und lyophilisiert; die letzte Lyophilysse wurde 12 h - 16 h durchgeführt.

4.2.10. Denaturierende Polyacrylamidgele

Lösungen für denaturierende 8%- bzw. 20%-(wt/vol) PAA-Gele (36) enthielten Acrylamid und N,N'-Methylenbisacrylamid im Verhältnis 19:1, 7 M Harnstoff, 100 mM Tris-Borat, pH 8,3, 2 mM EDTA, 0,07%(wt/vol) Ammoniumperoxodisulfat (39, 40). Nach Filtration und Entgasen der Lösungen wurde die Polymerisation mindestens 2 h vor Gebrauch der Gele durch 0,05%(vol/vol) TEMED gestartet; eine Verlängerung der Polymerisationsdauer von zunächst 2 h auf 12 h - 16 h bei Raumtemperatur erwies sich als vorteilhaft (37).

Lyophilisierte DNA aus 4.2.9 wurde in Wasser und in 70%(vol/vol) Formamid, 0,01%(wt/vol) Bromphenolblau, 0,01%(wt/vol) Xylen-cyanol, 1 mM EDTA, pH 7,4 (40) im Verhältnis 1:2 gelöst. Nach einer Vorelektrophorese von etwa 30 min wurde die 4 min bei 90°C denaturierte DNA mit 50 μ l-Glaskapillaren, deren Enden über einer Flamme zu langen, dünnen Spitzen ausgezogen waren, auf die Gele aufgetragen (40). Die Volumina wurden nach Möglichkeit so gewählt, daß pro Tasche etwa 10.000 Ipm in 3 μ l aufgetragen werden konnten. Die Elektrophorese von 20%-PAA-Gelen wurde bei 5 V/cm, von 8%-PAA-Gelen bei 4 V/cm durchgeführt.

4.2.11. Autoradiographie der Gele

Die Autoradiographie der mit einem Plastikfilm bedeckten Gele wurde bei -70°C durchgeführt (40), die Dauer der Exposition lag, wenn etwa 10.000 Ipm pro Bahn aufgetragen worden waren, bei 3 Tagen für 20%-Gele und bei 8 - 10 Tagen für 8%-Gele. In einigen Fällen wurden Verstärkerschirme benutzt (Ampli Universal, Cawo SE 6).

4.2.12. Sequenzauswertung

Da unserer Arbeitsgruppe keines der in der Literatur beschriebenen Programme für eine schnelle und einfache Auswertung von DNA-Sequenzen (49, 32, 54) zur Verfügung stand, wurden im Rahmen der vorliegenden Dissertation mehrere Programme in der Programmiersprache BASIC ("Beginner's All-purpose Symbolic Instruction Code") entwickelt und auf Magnetband gespeichert. Im einzelnen handelt es sich um folgende Programme, die durch ein Titelprogramm leicht auszuwählen sind:

1. "Speichern" ermöglicht die Speicherung von DNA-Sequenzen mit einer Länge bis zu 8,19 kb oder von Proteinen mit maximal 2730 Aminosäuren auf Magnetband. Die Markierung und Dimensionierung der dazu erforderlichen Datenspeicher erfolgt unter Programmkontrolle. Weiterhin können bereits gespeicherte DNA-Sequenzen in ihre Komplementärstränge übersetzt und um- oder neu gespeichert werden.
2. "Prüfen" untersucht, ob gespeicherte DNA-Sequenzen ausschließlich Symbole enthalten, die von anderen Programmen als Nucleotide erkannt werden (A, G, C, T, wahlweise auch R für Purine, Y für Pyrimidine und N für nicht definierte Basen).

Alle anderen Symbole werden aus der Datenkette entfernt. Dabei kann in jedem Fall zwischen einer ersatzlosen Deletion und einem Ersatz des falschen Symbols durch ein Leerzeichen als Platzhalter gewählt werden. Die korrigierten Sequenzen werden unter Programmkontrolle wieder gespeichert.

3. "Print" druckt gespeicherte Sequenzen oder Teile davon aus, wobei die Anzahl der pro Zeile gedruckten Nucleotide oder Aminosäuren und die Art der Darstellung bei DNA-Sequenzen (gespeicherter Strang oder dessen Gegenstrang, Einzel- oder Doppelstrang) vom Benutzer zu bestimmen sind.
4. "Plot" macht zeitraubendes Zeichnen von DNA-Sequenzen mit Schablone und Tusche überflüssig. DNA-Sequenzen können als Einzel- oder Doppelstrang dargestellt werden, die Zahl der pro Zeile aufzutragenden Nucleotide wird vom Benutzer bestimmt; die Größe der Zeichen richtet sich hierbei nach dem gewählten Format und der Zeilenlänge. Die ganze DNA-Sequenz oder auch einzelne Teile davon können während des Programmablaufs in Aminosäuresequenzen übersetzt werden; diese werden je nach Entscheidung des Benutzers im Ein- oder im Dreibuchstabencode unter die entsprechenden Codons gesetzt. (Siehe Abb. 4 als Beispiel.)
5. "Revision" erlaubt die Korrektur bereits gespeicherter Sequenzen. Es können Erweiterungen, Insertionen und Deletionen eingeführt und einzelne Elemente ausgetauscht werden. Die korrigierten Sequenzen werden unter Programmkontrolle um- oder neu gespeichert.
6. "Nachbarn" bestimmt die Basenzusammensetzung und die Häufigkeiten aller 16 Dinucleotide in gespeicherten DNA-Sequenzen, die vom Benutzer beliebig eingegrenzt werden können, und druckt die Ergebnisse aus. Wahlweise werden Einzel- oder Doppelstränge analysiert. Erwartungswerte für die Dinucleotidhäufigkeiten werden auf der Grundlage der Basenzusammensetzung der

analysierten DNA-Bereiche errechnet (56). Abweichungen von den Erwartungswerten werden als Differenzen und als Verhältnisse ausgedruckt. Weiterhin können auf Abruf die Positionen aller Dinucleotide im analysierten DNA-Bereich in histogramm-ähnlicher Form ausgedruckt werden:

Wenn ein bestimmtes Dinucleotid innerhalb von 16 Nucleotiden mehr als einmal und somit häufiger, als es einer groben Erwartung entspricht, vorkommt, werden die entsprechenden Positionen nebeneinander und nicht untereinander gedruckt. So werden DNA-Abschnitte, in denen bestimmte Dinucleotide gehäuft auftreten, leicht erkennbar.

7. "Suche" sucht in gespeicherten Sequenzen nach bestimmten, vom Benutzer abzufragenden Mustern und druckt, wenn solche in der Sequenz vorkommen, die entsprechenden Positionen aus.
8. "Homologie" sucht in gespeicherten Sequenzen nach wiederholten Teilsequenzen und nach Homologien mit einer, zwei oder drei anderen, vom Magnetband eingelesenen Sequenzen. Vom Benutzer ist zu bestimmen, ab welcher Länge Homologien als solche erkannt und ausgedruckt werden sollen.
9. "Symmetrie" sucht in gespeicherten DNA-Sequenzen nach Punkt- oder Spiegelsymmetrien (siehe auch 3.5); bis zu drei weitere DNA-Sequenzen können analog zum Programm "Homologie" in die Suche einbezogen werden. Vom Benutzer ist zu bestimmen, ab welcher Länge Symmetrien erkannt und ausgedruckt werden sollen.
10. "Translation" übersetzt gespeicherte DNA-Sequenzen in allen sechs möglichen Rastern in die entsprechenden Aminosäuresequenzen und druckt sie aus. Das Unterprogramm für die Translation ist in Abb. 10a aufgelistet.

11. "Protein" übersetzt einen vom Benutzer einzugrenzenden Teil einer gespeicherten DNA-Sequenz in die Aminosäuresequenz und speichert sie auf Magnetband ab.
12. "Triplett" erstellt eine Tabelle für die Häufigkeit der Verwendung der 64 möglichen Codons in einem vom Benutzer einzugrenzenden Bereich einer gespeicherten DNA-Sequenz. Für jedes Codon werden die Miyata-Indices berechnet (43) und in die Tabelle aufgenommen. Der Programmteil für die Translation und die Zählung der Codons ist in Abb. 10b aufgelistet.
13. "Spaltung" sucht die Erkennungssequenzen aller zur Zeit bekannten Restriktionsnucleasen in gespeicherten DNA-Sequenzen und druckt die entsprechenden Positionen aus.
14. "Random" generiert DNA-Sequenzen beliebiger Länge (bis maximal 8,19 kb) und speichert diese auf Magnetband ab. Diese Sequenzen sind Zufallssequenzen ähnlich, jedoch nicht absolut gleich, weil das verwendete System keine echte Zufallszahlenreihe produzieren kann.

Alle genannten Programme sind soweit als möglich gegen Bedienungs- und Eingabefehler abgesichert.

[illegible]

Abb. 10: Zwei mögliche Arten der Translation einer Nucleinsäuresequenz in eine Aminosäuresequenz durch ein Computerprogramm

In beiden Fällen wird eine Stringvariable (S\$ in a), F\$ in b) mit allen möglichen Übersetzungsergebnissen generiert. Die eigentliche Translation besteht in der Berechnung der Position des 1. Zeichens (S2 in a), B2 in b)) der betreffenden Aminosäure in den Stringvariablen. Von dieser Position an wird dann ein Substring (D\$ in a), I\$ in b)) mit einer Länge von 3 Zeichen, die betreffende Aminosäure, abgelesen. In a) wird die Position S2 durch Abfragen aller möglichen Codons über IF ... THEN-Kommandos errechnet, wobei bei jeder "NEIN"-Antwort der Zähler S1 um 1 erhöht wird. Aus dem Wert des Zählers S1 wird bei einer "JA"-Antwort S2 bestimmt (siehe Zeilen 7700 ff.). In b) wird die Position B2 bestimmt, indem die Position des jeweils zu translatierenden Codons (H\$) in einer Stringvariablen, die alle 64 möglichen Triplets enthält (D\$), gesucht wird. In F\$ entsprechen die Positionen der Aminosäuren denen der dafür codierenden Triplets in D\$. Der Vorteil von b) gegenüber a) liegt in einer höheren Translationsgeschwindigkeit, jedoch können bei a) auch Codons mit nicht eindeutig bestimmten Nucleotiden translatiert werden, soweit der degenerierte genetische Code dies zuläßt.

5. Zusammenfassung

Die Nucleotidsequenz des C_K-Gensegments und eines Teils des Introns zwischen diesem Bereich und den J-Gensegmenten wurde in DNA aus Leber- und Myelomgewebe von BALB/c-Mäusen bestimmt. Von Hamlyn et al. (24) und Max et al. (38) veröffentlichte Teile dieser Sequenz (Positionen 487-1018 bzw. 411-653 in Abb. 4) konnten bestätigt werden.

Es wurden zwischen DNA aus Leber- und aus Tumorgewebe keine Unterschiede hinsichtlich der Primärstruktur festgestellt; dies weist darauf hin, daß im sequenzierten DNA-Bereich keine somatischen Mutationen stattgefunden haben.

Die Poly-(A)-Anheftungsstelle ist wie bei anderen Genen durch den Vergleich der Nucleotidfolge genomischer DNA mit der korrespondierenden mRNA-Sequenz nicht genau festzulegen.

Die Computerauswertung der ermittelten Sequenz erfolgte unter verschiedenen Aspekten: Verschiedene Segmente weichen in der Basenzusammensetzung ihrer Einzelstränge zum Teil stark voneinander ab. Der T-Reichtum des 3'-Teils des Introns steht im Einklang mit einer Hypothese von Bina et al. (4) über den Ablauf des RNA-Spleißens. Im C_K-Gensegment kommt das Dinucleotid CG signifikant selten vor, CA und TG dagegen werden signifikant häufig gefunden. Dieser Sachverhalt wird unter Berücksichtigung der Codon-Verwendung im Hinblick auf die Methylierung-Mutations-Theorie von Bird (5) diskutiert. Die von anderen Autoren (43) geäußerte Vermutung, daß sich aus der Codon-Verwendung im C_K-Gensegment Rückschlüsse auf die Ausbildung einer Sekundärstruktur in der mRNA ziehen lassen, konnte anhand der

Sequenzdaten nicht erhärtet werden; jedoch ergibt sich aus der Nucleotidfolge des sequenzierten Teils des Introns eine mögliche Sekundärstruktur im mRNA-Vorläufer.

5. Literaturverzeichnis

1. Abelson, J. (1979) Ann. Rev. Biochem. 48, 1035-1069
2. Altenburger, W., Steinmetz, M. und Zachau, H.G. (1980) Nature 287, 603-607
3. Benoist, C., O'Hare, K., Breathnach, R. und Chambon, P. (1980) Nucl. Acids Res. 8, 127-142
4. Bina, M., Feldmann, R.J. und Deeley, R.G. (1980) Proc. Natl. Acad. Sci. USA 77, 1278-1282
5. Bird, A.P. (1980) Nucl. Acids Res. 8, 1499-1504
6. Blattner, F.R., Williams, B.G., Blechl A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. und Smithies, O. (1977) Science 196, 161-169
7. Boyer, H.W., Chow, L.T., Dugaiczky, A., Hedgpeth, J. und Goodman, H.M. (1973) Nature New Biol. 244, 40-43
8. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. und Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857
9. Burstein, Y. und Schechter, I. (1978) Biochemistry 17, 2392-2400
10. Burton, K. (1956) Biochem. J. 62, 315-323
11. Burton, K. (1967) in Methods in Enzymology, Vol. 12, Part A (Hrsg. Grossman, L. und Moldave, K.), pp.222-224 (Academic Press, New York)
12. Cashmore, A.R. und Petersen, G.B. (1978) Nucl. Acids Res. 5, 2485-2491
13. Cavalier-Smith, T. (1976) Nature 262, 255-256
14. Chaconas G. und van de Sande, J.H. (1980) in Methods in Enzymology, Vol. 65, Part I (Hrsg. Grossman L. und Moldave, K.), pp.75-85 (Academic Press, New York)
15. Doskocil, J. und Sorm, F. (1962) Biochim. Biophys. Acta 55, 953-959
16. Dreyer, W.J. und Bennett, J.C. (1965) Proc. Natl. Acad. Sci. USA 54, 864-869

17. Edelman, G.M. und Benacerraf, B. (1962)
Proc. Natl. Acad. Sci. USA 48, 1035-1042
18. Edelman, G.M. und Gally, J.A. (1962)
J. Exptl. Med. 116, 207-227
19. Freedman, M.H. und Sela, M. (1966) J. Biol.
Chem. 241, 2383-2396
20. Gilbert, W. (1978) Nature 271, 501
21. Glynn, I.M. und Chappell, J.B. (1964)
Biochem. J. 90, 147-149
22. Gray, C.P., Sommer, R., Polke, C., Beck, E.
und Schaller, H. (1977) Proc. Natl. Acad. Sci.
USA 75, 50-53
23. Haber, E. (1964) Proc. Natl. Acad. Sci. USA 52,
1099-1106
24. Hamlyn, P.H., Brownlee, G.G., Cheng, C.-C.,
Gait, M.J. und Milstein, C. (1978) Cell 15,
1067-1075
25. Hasegawa, M., Yasunaga, T. und Miyata, T. (1979)
Nucl. Acids Res. 7, 2073-2079
26. Hilschmann, N. und Craig, L.C. (1965) Proc. Natl.
Acad. Sci. USA 53, 1403-1409
27. Hörz, W., Hess, I. und Zachau, H.G. (1974) Eur. J.
Biochem. 45, 501-512
28. Kochetkov, N.K. und Budovskii, E.I. (1972) in
Organic Chemistry of Nucleic Acids, Part B,
pp.381-385 (Plenum Press, London und New York)
29. pp.397-400 in (28)
30. pp.401-408 in (28)
31. pp.504-507 in (28)
32. Korn, L.J., Queen, C.L. und Wegman, M.N. (1977)
Proc. Natl. Acad. Sci. USA 74, 4401-4405
33. Lawley, P.D. und Brookes, P. (1963) Biochem. J. 89,
127-138
34. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L.
und Steitz, J.A. (1980) Nature 283, 220-224

35. Mach, B., Rougeon, F., Longacre, S. und Aellen, M.-F. (1977) in Molecular Cloning of Recombinant DNA, Miami Winter Symposia, Vol. 13 (Hrsg. Scott, W.A. und Werner, R.) pp.219-235 (Academic Press, New York)
36. Maniatis, T., Jeffrey, A. und van deSande, H. (1975) Biochemistry 14, 3787-3794
37. Mantei, N., Schwarzstein, M., Streuli, M., Panem, S., Nagata, S. und Weissmann, C. (1980) Gene 10, 1-10
38. Max, E.E., Seidman, J.G. und Leder, P. (1979) Proc. Natl. Acad. Sci. USA 76, 3450-3454
39. Maxam, A.M. und Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
40. Maxam, A.M. und Gilbert, W. (1980) in Methods in Enzymology, Vol. 65, Part I (Hrsg. Grossman L. und Moldave, K.), pp.499-560 (Academic Press, New York)
41. Miller, J.R., Cartwright, E.M., Brownlee, G.G., Fedoroff, N.V. und Brown, D.D. (1978) Cell 13, 717-725
42. Milstein, C., Brownlee, G.G., Harrison, T.M. und Mathews, M.B. (1972) Nature New Biol. 239, 117-120
43. Miyata, T., Hayashida, H., Yasunaga, T. und Hasegawa, M. (1979) Nucl. Acids Res. 7, 2431-2438
44. Ohmori, H., Tomizawa, J. und Maxam, A.M. (1978) Nucl. Acids Res. 5, 1479-1485
45. Peacock, A.C. und Dingman, C.W. (1967) Biochemistry 6, 1818-1827
46. Pech, M. (1980) Dissertation, Universität München
47. Pech, M., Streeck, R.E. und Zachau, H.G. (1979) Cell 18, 883-893
48. Pfeiffer, W., Hörz, W., Igo-Kemenes, T. und Zachau, H.G. (1975) Nature 258, 450-452
49. Pipas, J.M. und McMahon, J.E. (1975) Proc. Natl. Acad. Sci. USA 72, 2017-2021
50. Potter, M. (1972) Physiol. Rev. 52, 631-719
51. Potter, M. und Robertson-Boyce, C. (1962) Nature 193, 1086-1087

52. Proudfoot, N.J. und Brownlee, G.G. (1976)
Nature 263, 211-214
53. Putnam, F.W., Easley, C.W. und Helling, J.W.
(1963) Biochim. Biophys. Acta 78, 231-233
54. Queen, C.L. und Korn, L.J. (1980) in Methods in
Enzymology, Vol. 65, Part I (Hrsg. Grossman, L.
und Moldave, K.), pp. 595-609 (Academic Press,
New York)
55. Richardson, C.C. (1965) Proc. Natl. Acad. Sci.
USA 54, 158-165
56. Russell, G.J. und Subak-Sharpe, J.H. (1977)
Nature 266, 533-536
57. Sakano, H., Hüppi, K., Heinrich, G. und Tonegawa,
S. (1979) Nature 280, 288-294
58. Salser, W. (1977) Cold Spring Harbor Symp. Quant.
Biol. 42, 985-1002
59. Sanger, F. und Coulson, A.R. (1978) FEBS Letters
87, 107-110
60. Scarano, E., Iaccarino, M., Grippo, P. und
Parisi, E. (1967) Proc. Natl. Acad. Sci. USA
57, 1394-1400
61. Schiff, C. und Fougereau, M. (1975) Eur. J.
Biochem. 59, 525-537
62. Seidman, J.G., Leder, A., Nau, M., Norman, B.
und Leder, P. (1978) Science 202, 11-17
63. Seidman, J.G., Max, E.E. und Leder, P. (1979)
Nature 280, 370-375
64. Setlow, P. (1976) in Handbook of Biochemistry
and Molecular Biology - Nucleic Acids, Vol. 2
(Hrsg. Fasman, G.D.), pp.312-318 (Cleveland,
Ohio: CRC Press)
65. Shapiro, H.S. (1967) in Methods in Enzymology,
Vol. 12, Part A (Hrsg. Grossman, L. und Moldave,
K.), pp.212-218 (Academic Press, New York)
66. Shapiro, H.S. (1976) in Handbook of Biochemistry
and Molecular Biology - Nucleic Acids, Vol. 2
(Hrsg. Fasman, G.D.), pp.241-311 (Cleveland, Ohio:
CRC Press)

67. Smith, G.P., Hood, L. und Fitch, W.M. (1971) Ann. Rev. Biochem. 40, 969-1012
68. Smith, H.O. und Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2398
69. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517
70. Steinmetz, M. und Zachau, H.G. (1980) Nucl. Acids Res. 8, 1693-1707
71. Steinmetz, M., Zachau, H.G. und Mach, B. (1979) Nucl. Acids Res. 6, 3213-3229
72. Sussenbach, J.S., Steenbergh, P.H., Rost, J.A., van Leeuwen, W.J. und van Embden, J.D.A. (1978) Nucl. Acids Res. 5, 1153-1163
73. Swartz, M.N., Trautner, T.A. und Kornberg, A. (1962) J. Biol. Chem. 237, 1961-1967
74. Thimmappaya, B. und Shenk, T. (1979) J. Virol. 30, 668-673
75. Thomas, C.A. jr., Feyeritz, R.E., Wilson, D.A., Darcis, F.V., Lee, C.S., Bick, M.D., Huang, H.L. und Zimm, B.H. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 353-370
76. Titani, K., Whitley, E. jr., Avogardo, L. und Putnam, F.W. (1965) Science 149, 1090-1092
77. Volckaert, G., Feunteun, J., Crawford, L.V., Berg, P. und Fiers, W. (1979) J. Virol. 30, 674-682
78. Waring, M.J. (1965) J. Mol. Biol. 13, 269-282
79. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. und Hood, L. (1978) Nature 276, 785-790
80. Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J. und Hood, L. (1980) Nature 283, 497-499
81. Weiss, B., Live, T.R. und Richardson, C.C. (1968) J. Biol. Chem. 243, 4530-4542
82. Wilson, D.A. und Thomas, C.A. jr. (1974) J. Mol. Biol. 84, 115-144
83. Whitney, P.L. und Tanford, C. (1965) Proc. Natl. Acad. Sci. USA 53, 524-532
84. Wu, T.T. und Kabat, E.A. (1970) J. Exp. Med. 132, 211-250

7. Abkürzungsverzeichnis

AluI	Restriktionsnuclease aus <i>Arthrobacter luteus</i>
BALB/c	Inzucht-Mäusestamm
bp	Basenpaar
BspRI	Restriktionsnuclease aus <i>Bacillus sphaericus</i> R
C	konstant
cDNA	komplementäre DNA
EcoRI	Restriktionsnuclease aus <i>Escherichia coli</i> RY13
EcoRII	Restriktionsnuclease aus <i>Escherichia coli</i> R245
H-Ketten	schwere Immunglobulin-Ketten
HinfI	Restriktionsnuclease aus <i>Haemophilus influenzae</i> R _f
HpaI	Restriktionsnuclease aus <i>Haemophilus parainfluenzae</i>
Intron	intragenic region
Ipm	Impulse pro Minute
J-Gensegmente	für die Aminosäuren 96-108 von κ -Ketten codierende DNA
κ	Immunglobulin-Typ
kb	1000 Basenpaare
λ	Immunglobulin-Typ
L-Gensegmente	für hydrophobe Präsequenzen codierende DNA
L-Ketten	leichte Immunglobulin-Ketten
m ⁵ C	5-Methyl-cytidin
MOPC	mineral oil-induced plasmocytoma
PAA	Polyacrylamid
Sau3A	Restriktionsnuclease aus <i>Staphylococcus aureus</i> 3A
Sau 96I	Restriktionsnuclease aus <i>Staphylococcus aureus</i> PS96
TEMED	N,N,N',N',-Tetramethyläthylendiamin
U	Enzymaktivitätseinheit
3'UT	3'-nicht-translatierte Region in mRNAs
V	variabel
vol	Volumen
wt	Gewicht

Alle übrigen, hier nicht aufgeführten Abkürzungen entsprechen den "instructions to authors" des European Journal of Biochemistry.

Lebenslauf

Name: Peter Sebastian Neumaier

Geboren: 11.Juni 1955 in München

Eltern: Sebastian Neumaier, geb. 17.Mai 1924
Hermine Neumaier, geb. 29.Juni 1929

Schulen: 1961 - 1966 Volksschule in München
1966 - 1975 Ludwigsgymnasium
in München

Allgemeine
Hochschulreife: 21.Juni 1975

Wehrdienst: 1975 - 1977 in Murnau

Studium der
Humanmedizin: 1977 - 1983 Ludwig Maximilians-
Universität München

Dissertation: 1979 - 1980 am Institut für Physiologische
Chemie, Physikalische Biochemie
und Zellbiologie der Universi-
tät München

Ärztliche Prüfung: 7.November 1983

Approbation: 18.November 1983

